The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

The tangled history of olfaction in African mole-rats, Bathyergidae: insights from Olfactory Receptor genes

Sofia Stathopoulos

University of Cape Town

Supervisors:
Dr Colleen O'Ryan, Department of Molecular and Cell Biology
Dr Jacqueline M Bishop, Department of Zoology

Thesis presented for the degree of

DOCTOR OF PHILOSOPHY

In the Department of Molecular and Cell Biology

Faculty of Science

UNIVERSITY OF CAPE TOWN

August 2011

DECLARATION

This thesis reports the results of the original research I conducted under the auspices of the Department of Molecular and Cell Biology in the Faculty of Science at the University of Cape Town, between 2008 and 2011. All the assistance that I received has been acknowledged. This work has not been submitted for a degree at any other university.

University of Cape Town

_____

Sofia Stathopoulos

Abstract

Stathopoulos S (2011) The tangled history of olfaction in African mole-rats, Bathyergidae: insights from Olfactory Receptor genes. *PhD Thesis*, University of Cape Town, South Africa.

Vertebrate olfactory receptors (ORs) belong to a multi-gene family with >1500 genes known in mice. The polymorphism of these genes reflects the diversity of odorants detected and discriminated, and is correlated with the olfactory sensitivity of a species. The gene family evolves rapidly via a 'birth and death' model, where selection shapes gene diversity. African mole-rats (family Bathyergidae) are subterranean rodents endemic to sub-Saharan Africa and display varying levels of sociality. Life underground has imposed unusual constraints on social interactions and successful foraging, resulting in a suite of physiological, sensory and behavioural adaptations. Given the selective pressures of life underground, enhanced olfaction is considered fundamental to the evolutionary success of the Bathyergidae; enhanced diversity at OR genes is therefore predicted to characterise the bathyergid OR gene family. This thesis reports the first assessment of OR variation in Bathyergidae, and therefore, the first for a family of subterranean mammals. Using a PCR-sequencing approach, 178 unique OR sequences, corresponding to 119 unique OR genes are characterised from 14 mole-rat species. Bathyergidae OR genes are classified using sequence similarity and phylogenetic comparison with more than 50 mammalian OR subgenomes. Using a combination of classic tests and tree-based methods, the mechanisms of molecular evolution across the mole-rat OR gene tree are explored. Four well-supported clades emerge in the gene phylogeny, with varying signals of selection: from no apparent selection, to positive selection, to purifying selection. This is symptomatic of the diverse olfactory recognition properties of OR genes, and is suggestive of a fundamental role of ORs in Bathyergidae. The relative contributions of sociality and environmental niche-specialisation in driving natural selection on ORs are also tested. Both factors are explored as potential drivers of bathyergid OR evolution within a modern bioinformatic framework, contributing to our understanding of OR diversity across different mammalian ecotypes.

University of Cape Town

Acknowledgments

I would like to thank my supervisors, Dr Jacqueline M Bishop and Dr Colleen O'Ryan, for giving me the unique opportunity to carry out this exciting research. I am particularly grateful to Dr Jacqueline M Bishop for her precious intellectual guidance and for sharing her expert knowledge in the fields of molecular evolution and the ecology of this interesting subterranean family. Dr Colleen O'Ryan provided me with immense intellectual and psychological support throughout this PhD, sharing her expertise and helping me to focus and achieve this sought-after goal.

Prof Nigel C Bennett (Department of Zoology & Entomology, University of Pretoria, South Africa) is thanked for supplying all the bathyergid samples that made this research possible.

The South African National Research Fund, NRF, provided financial support for the work detailed in this thesis.

I am especially grateful to my husband, Renaud Gaujoux, for his invaluable help in all the bioinformatic aspects of this thesis, as well as his love and support throughout the arduous PhD journey that we embarked together.

I would also like to thank all my lab colleagues, for the technical advice and the cheerful conversations in times of need. A very special thanks goes to my friend Natashia Muna, with whom I shared endless conversations on both science and motherhood, as well as delicious coffees, teas, baked snacks and lunches, which helped to brighten-up my darkest days in the lab.

I am especially grateful to my family for supporting me during this PhD, and in particular to my daughters, Paola and Dafne, for reminding me every day that everything is possible.

University of Cape Town

Table of Contents

University of Cape Town

University of Cape Town

Chapter 1: General Introduction

1.1 The nature of Olfaction

'*The air came laden with the fragrance it caught upon its way, and the bees, upborne upon its scented breath, hummed forth their drowsy satisfaction as they floated by.*'

Charles Dickens (1841) *The Old Curiosity Shop*

University of Cape Town

Olfaction is a vitally important sense for all living organisms. As humans - with our highly developed visual and auditory senses - we often fail to appreciate the significance of a well developed sensitivity to the chemical world that surrounds us. Indeed, for most animals, chemical cues represent the main source of information about their social and physical environment, and their dominant means of communication. Even the simplest life forms, from bacteria to protozoans, are sensitive to 'chemical irritation' (e.g. Van Houten 1994, Taga and Bassler 2003). A predisposition of all living cells to react to chemical stimuli presumably played an important role in the evolution of specific receptor molecules, leading the way for the development of simple sensory organs capable of detecting chemical information, and ultimately the complex olfactory systems that can be found in all vertebrate species today (Ache and Young 2005).

Olfactory cues come in many different forms. Environmental odours direct animals to desirable locations, e.g. to find food or water, and warn of potential dangers such as rotten food or fires. 'Allelochemic' substances are odours secreted by individuals from one species, that are perceived by another species (Whittaker and Feeny 1971). These substances regulate more complex behaviours e.g. prey localisation, predator deterrence and avoidance, territorial marking, and pollination (Langley 1988, Mathis and Vincent 2000, Monclus et al. 2009, Raguso and Willis 2002). Environmental and allelochemic odorants are often referred to as

'general' odorants, as opposed to odours of conspecific origin, commonly known as 'pheromones'. Pheromones were first defined by Karlson and Lüscher (1959) as 'biological compounds that are secreted and have a defined physiological or behavioural effect on an individual of the same species'. Pheromones modulate complex behaviours in both vertebrates and invertebrates, such as social, aggressive, reproductive and sexual behaviours among conspecifics (Wyatt 2003). Inter-male aggression, male sexual preference, puberty acceleration, maternal aggression, and pregnancy block are examples of typical pheromone-induced behaviours (Del Punta et al. 2002, Halpern et al. 2003, Leypold et al. 2002, Stowers et al. 2002, Norlin et al. 2003).

For air-breathing organisms, olfactant molecules are volatile compounds with a molecular weight of lower than 300 Da circa, that are, by definition, perceived as 'odorous' by the olfactory system (Touhara and Vosshall 2009). Odorant molecules are water-soluble for aquatic animals, whereas they are typically small hydrophobic organic molecules for terrestrial animals (Mombaerts 1999a). Pheromones, on the other hand, can be non-volatile substances with a higher molecular weight than general odorants, including relatively large organic compounds, peptides and proteins (Touhara and Vosshall 2009). Pheromones are not necessarily odorous, as long as the chemical signals conveyed are perceived by conspecifics (Touhara and Vosshall 2009).

One may ask, how many odorants can be detected and distinguished by humans? Given that a large proportion of volatile compounds have a discernable odour, the number of odorous chemicals present in the terrestrial environment could be of the range of hundreds of thousands, if not millions (Mombaerts 2004). Nevertheless, most olfaction research is based on a few dozen odorants from a standard set of ~500 chemicals, due to the limited availability of commercial odorants (Haddad et al. 2008). In humans, the most often cited statistic of being sensitive to ~10,000 detectable odorants does not account for the difference between 'detection' and 'identification'. Oenologists and perfume creators, supposedly the best 'noses' in the Wine and Fragrance industries, are still capable of distinguishing only a few thousand odours by name (Sarafoleanou et al. 2009).

Highly developed olfaction and odour discrimination are associated with a broad range of fitness-related behaviours in mammals, from foraging and danger avoidance, to the complex behavioural processes of individual recognition, mate choice and maternal care, all of which

University of Cape Town

are based on chemosensory communication (Swaisgood et al. 1999, Firestein 2001, Brennan & Kendrick 2006). The advantages of olfactory communication are numerous. Secreted signals persist in the environment after the signaller has dispersed (Müller-Schwartze 2006) and thus can be energetically efficient to produce (Ginzel 2010, but see Gosling et al. 2000). Information gathering and territorial defence are achieved without physical contact, theoretically limiting the risk to individuals (Muller-Schwartze 2006). For example, house mice (*Mus domesticus*) use persistent scent marks in the form of small urine smears for territorial advertisement, individual recognition and mate choice, in their complex social networks (Hurst and Beynon 2004). Dominant males are the main signallers within mice societies, and concentrate scent marking in areas that delimit their territory, as well as around valuable resources (Hurst and Beynon 2004). Olfactory cues are investigated equally by all the members of a mice society, by direct contact with the scent mark (Hurst and Beynon 2004). Such informative scent marks that are destined to conspecifics, can however be attractive to 'olfactorily-hunting' predators. Hughes et al. (2010a) found that mice predators e.g. cats (*Felis catus*) and snakes (*Pseudonaja textilis*) are strongly attracted to mouse-scented locations, increasing the predation risks for signal 'receivers'. Interestingly, this results in an observed trade-off between the benefits of receiving social signals and the perceived predation risk (Hughes et al. 2009, 2010b). For example, mice limit their use of 'low value' social signals when the predation risk is perceived as high, while they do not refrain from investigating 'high value' social signals (i.e. the smell of an intruder) under the same conditions (Hughes et al. 2009). Olfactory communication has the additional advantage of occuring independently of light for crepuscular, nocturnal or subterranean species. Prosimian primate species, for example, which are often nocturnal and live in dense vegetation, use olfactory communication to find mates, advertise reproductive and dominance status, coordinate mating and maintain territories (Charles-Dominique 1977, Schilling 1979, Dixson 1998, Kappeler 1998). In the nocturnal pigmy loris (*Nycticebus pygmaeus*), in particular, females choose mates based on their ability to countermark scent marks of competitor males (Fisher et al. 2003).

Indeed, impaired olfaction represents a significant threat to most species. Recent studies on the effects of anthropogenic ocean acidification reveal that elevated levels of atmospheric carbon dioxide (CO_2) causing a decrease in water pH (The Royal Society 2005, Fabry et al. 2008), result in disrupted olfactory abilities in fish (Munday et al. 2009, Dixson et al. 2010). The ecological success of many coastal marine species depends on homing abilities at the larval stage, i.e. on larvae finding a suitable adult habitat at the end of an offshore dispersive

University of Cape Town

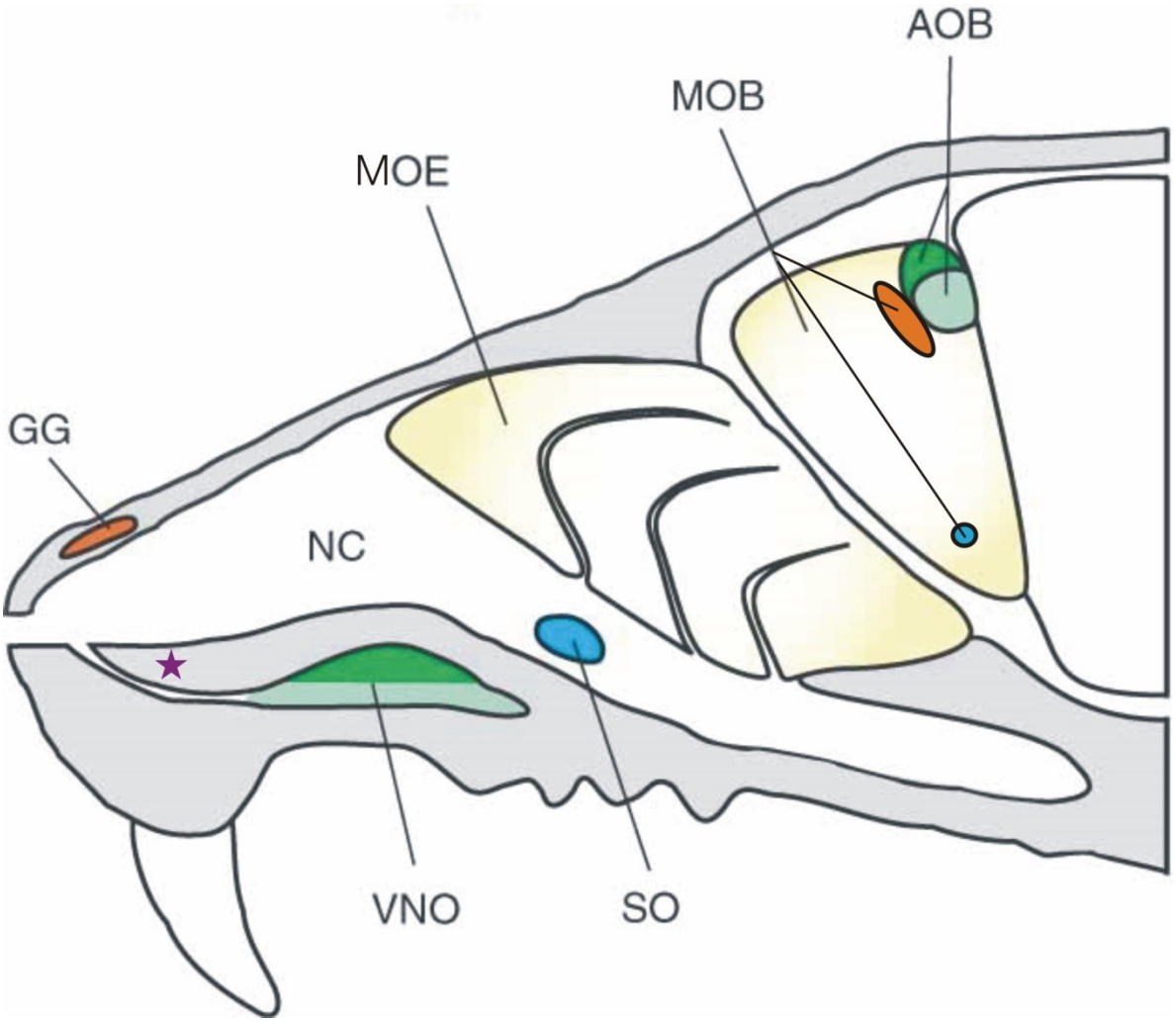
stage. At a pH of 7.8, predicted to be attained in ~2100 based on a '*business as usual CO_2 emissions trajectory*' (The Royal Society 2005, Caldeira and Wickett 2005, Orr et al. 2005), larval clownfish become strongly attracted to olfactory cues that they normally avoid (Munday et al. 2009). A further decrease of 0.2 pH units (expected to occur later next century, The Royal Society 2005) causes clownfish to stop responding to olfactory cues altogether (Munday et al. 2009), disrupting the natural homing behaviour of the species. Furthermore, clownfish larvae innately use olfactory cues to avoid predators, a behavioural trait that is retained though to adulthood (Dixson et al. 2010). When reared at a pH of 7.8, clownfish larvae become attracted to the smell of predators and lose the ability to distinguish between predatory and non-predatory fish (Dixson et al. 2010). The impairment of olfactory abilities directly affects the species' fitness by increasing mortality, due to a higher predation risk; a disastrous scenario in a future of acidified ocean conditions.

1.1.1 Olfactory mechanisms

Odour detection in mammals is accomplished via a complex olfactory system composed of distinct chemosensory subsystems; these include the Main Olfactory Epithelium, MOE, the Vomeronasal Organ, VNO, the Septal Organ of Masera, SO, and the Grueneberg Ganglion, GG (Breer et al. 2006, Spehr et al. 2006b, Ma 2007, Munger et al. 2009). Projections of sensory neurons connect these olfactory organs to specific areas of the brain that are dedicated to olfactory signal processing: the main olfactory bulb, MOB, and the accessory olfactory bulb, AOB (Figure 1.1, Tirindelli et al. 2009).

University of Cape Town

Figure 1.1 Rodent olfactory anatomy (adapted from Ferrero and Liberles 2010). Nasal compartments are labelled as follows: main olfactory epithelium, MOE, vomeronasal organ, VNO, septal organ, SO, and Grueneberg ganglion, GG. Odorants enter the nasal cavity (NC) during inhalation, and access the MOE, or the VNO by being pumped through the small duct indicated with a star. Sensory neurons of the MOE (beige), SO (blue), and GG (orange) project to areas of the same colour in the main olfactory bulb (MOB); sensory neurons of the apical and basal VNO (dark and light green, respectively) project to the corresponding areas of the accessory olfactory bulb (AOB).



Among these nasal compartments, it is generally thought that the MOE and VNO are the two main olfactory organs, responsible for the detection of most odorant compounds and pheromones (Touhara and Vosshall 2009). The MOE is present in all vertebrates except some cetaceans (Kishida et al. 2007), and a functional VNO is found in most tetrapods, including snakes and lizards (Haplern and Kubie 1980, Graves and Halpern 1990), as well as many mammals, e.g. mice, rats, horse, cattle, dogs, cats (Bellringer et al. 1980, Vaccarezza et al. 1981, Taniguchi and Mikami 1985, Salazar et al. 1992, Eccles 1982). In the MOE and VNO, odour detection is achieved via highly specialized chemosensory cells, the olfactory sensory

neurons (SNs). It had long been hypothesised that SNs in the MOE responded to general odorants, while the VNO was thought to be specialised for the detection of pheromones (Sam et al. 2001, Baxi et al. 2006, Brennan & Zufall 2006, Spehr et al. 2006b). However, it is now well established that the MOE and VNO have overlapping roles: MOE sensory neurons are activated by many olfactants that are effective in stimulating VNO neurons; VNO neurons, in turn, can respond to both non-volatile and volatile odorants that are not obviously pheromonal stimuli (Xu et al. 2005, Baxi et al. 2006, Brennan & Zufall 2006, Spehr et al. 2006b, Jakupovic et al. 2008). In line with this, species that lack a functional VNO, e.g. humans, are thought to achieve pheromonal communication via the MOE (Wang et al. 2007).

1.1.2 Olfactory receptors

The responsiveness of SNs to distinct odorants is determined by specialized receptors in their chemosensory membranes. Twenty years ago, Linda Buck and Richard Axel (1991) characterised the first set of Olfactory Receptor genes and were thus the first to elucidate the complex mechanisms of olfaction. For their pioneering work, Buck and Axel were awarded the Nobel Prize in Medicine or Physiology in 2004.

To unravel the mechanisms of olfaction, Buck and Axel's (1991) experimental approach was based on three assumptions. Firstly, that ORs were likely guanine-nucleotide protein (G-protein), coupled receptors (GPCRs), which characteristically have seven transmembrane (TM) domains. Secondly, that olfaction requires a large repertoire of receptors to match the diversity in chemical structure of known olfactants, thus ORs were likely to be members of a multigene family. Thirdly, because of their function, ORs were predicted to be selectively expressed in olfactory SNs.

The first and most significant assumption was based on biochemical evidence that G-proteins are involved in olfactory signal transduction. Exposure of olfactory SN cilia to odorants causes adenylate cyclase activation, which in turn, depends on the presence of guanosine triphosphate, GTP (Pace et al. 1985, Sklar et al. 1986). Thus, activation involves a GTP-binding protein such as $G_{\alpha olf}$, an olfactory SN-specific GPCR involved in odorant signal transduction which had been described prior to Buck and Axel's seminal work (Jones and Reed 1989). Proteins in the GPCR family typically have seven TM domains and are characterised by conserved regions, facilitating the design of oligonucleotide primers to target

University of Cape Town

these conserved regions in a polymerase chain reaction (PCR). Buck and Axel (1991) designed a series of degenerate primers which were used in all possible combinations on cDNA extracted from rat olfactory epithelium. Because the olfactory epithelium undoubtedly contained other 7TM GPCR proteins, Buck and Axel (1991) narrowed their search with the second criterion. To verify that a multigene family was being targeted, restriction enzyme digests were performed on PCR products. When the molecular weight of digested PCR products corresponded to multiples of the predicted molecular weight of undigested products, it was assumed that multiple genes had been amplified. A particular pair of primers, named A4 and B6, produced the majority of PCR products that satisfied the second criterion. To ensure that members of this multigene family were selectively expressed in the rat olfactory epithelium, a Northern Blot analysis was done. Genes that satisfied all three criteria were then cloned and sequenced to infer genetic identity. Following this elegant experimental procedure, Buck and Axel (1991) isolated and characterised 18 novel genes. All the genes displayed known motifs of GPCRs, but also shared some unusual features and were therefore representative of a new family of receptors - the Olfactory Receptor multigene family.

In the past two decades, extensive research on OR gene structure has corroborated Buck and Axel's original findings (1991), and an unexpectedly large repertoire of olfactory receptors has emerged. Olfactory receptors represent the molecular basis for the vast capacity of the olfactory system to detect and discriminate multitudes of olfactants. Based on their structure and expression patterns, olfactory receptors belong to a number of different receptor families which include the odorant receptors (ORs) initially described by Buck and Axel (1991), vomeronasal receptors (VRs, divided in V1Rs and V2Rs, Dulac and Axel 1995, Herrada and Dulac 1997, Matsunami and Buck 1997, Ryba and Tirindelli 1997), trace amine-associated receptors (TAARs, Borowsky et al. 2001, Lindemann et al. 2005), formyl peptide receptors (FPRs, Boulay et al. 1990, Riviere et al. 2009), and the guanylyl cyclases type D (GC-D, Fülle et al. 1995, Juilfs et al. 1997). Spatial expression patterns of these receptor families are shown in Figure 1.2.

University of Cape Town

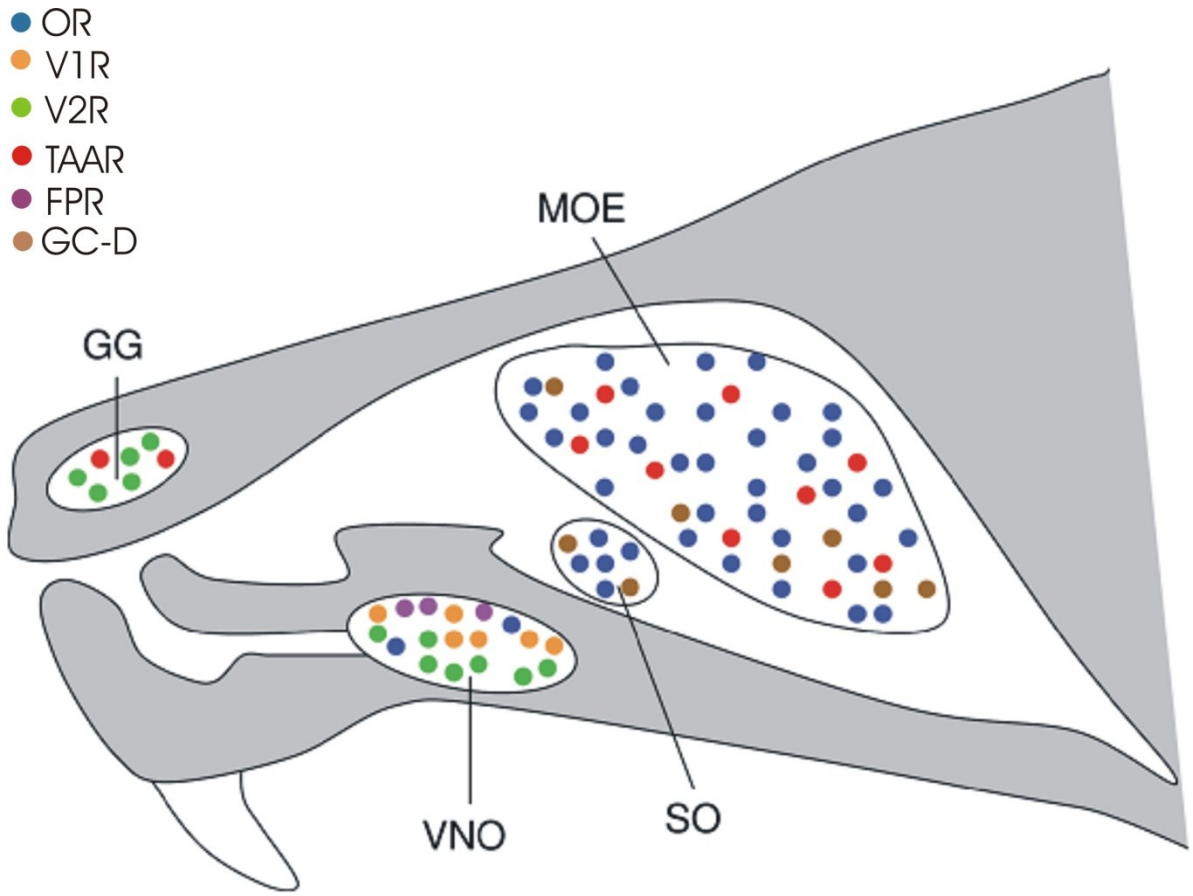


Figure 1.2 General expression patterns of different receptor families in the nasal compartments of rodents (adapted from Fleischer et al. 2009). Nasal compartments: MOE, main olfactory epithelium; VNO, vomeronasal organ; GG, Grueneberg Ganglion; SO, Septal organ. Receptor types: OR, olfactory (odorant) receptor; V1R and V2R, vomeronasal receptors type 1 and 2, respectively; TAAR, trace amine-associated receptor; FPR, formyl peptide receptor; GC-D, guanylyl cyclases type D.

University of

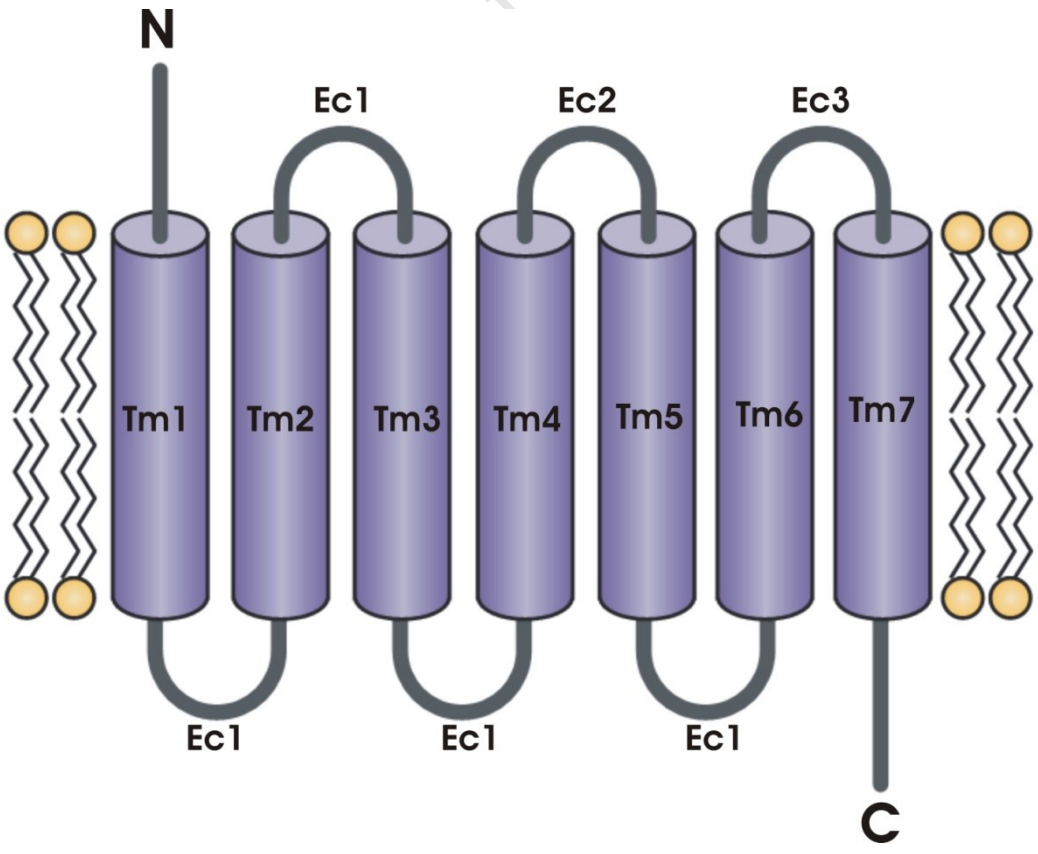
Of the aforementioned receptor families, ORs and VRs are the most important and best characterised receptors in vertebrates. This thesis focuses on vertebrate ORs, thus their main features will be reviewed in this chapter. A brief comparison will be made with VRs and TAARs, while FPRs and guanylyl cyclases GC-D will not be considered due to their poorly characterised olfactory function (Fleischer 2009).

Vertebrate odorant receptors, ORs

This section presents a concise overview of the structural, functional, genetic, genomic and expression characteristics of vertebrate odorant receptors, ORs, as well as a short description of the receptor families that complement the role of ORs in olfaction. A more exhaustive description of OR genetic and evolutionary features is presented in the chapters that follow.

Vertebrate 'odorant' receptors, generally referred to as 'olfactory receptors', ORs, are GPCR proteins that have seven TM α- helical hydrophobic regions, three intracellular (IC) and three extracellular (EC) loops, as well as an extracellular N-terminal and an intracellular C-terminal domain (Figure 1.3). Based on their primary structure GPCR proteins are characterised into three classes, A, B or C, with very little shared sequence homology between them (Jacoby et al. 2006); in this classification, ORs belong to class A GPCRs due to their domain organisation. Class A receptors account for ~85% of GPCR genes and include several opsins e.g. rhodopsin, a pigment expressed in the retina of the eye; a number of peptide receptors eg. chemokine receptors, various biogenic amine receptors e.g. adrenergic or dopamine receptors, as well as hormone protein receptors e.g. follicle-stimulating hormone (FSH) and luteinising hormone (LH) receptors, and many others (Gether 2000).

Figure 1.3 Typical GPCR class A structure (adapted from Nei et al. 2008). Tm1-7, transmembrane domains 1-7; Ec1-3, extracellular domains 1-3; Ic1-3, intracellular domains 1-3; N, amino-terminus (extracellular); C, carboxyl-terminus (intracellular). ORs, V1Rs and TAARS share this membrane topology.



f Cape Town

The average length of OR proteins is of about 320 ± 25 amino acid residues, with differences in size resulting from variable N- and C- terminal stretches. Specific amino-acid motifs are used to distinguish ORs from other GPCRs. These conserved motifs include a LHTPMY motif in the first intracellular loop (IC1), the most characteristic MAYDRYVAIC motif within TM domain 3 (TM3), an SY motif in TM5, FSTCSSH in TM6 and PMLNPF in TM7 (Appendix I.1; Zozulya et al. 2001, Fleischer et al. 2009). Conserved OR motifs may differ slightly across species, but they are generally used to identify OR genes from all the vertebrate genomes studied to date (Fleischer et al. 2009).

OR ligand binding and signalling

When Buck and Axel first described OR genes (1991), they noted that the most variable regions of the putative receptor proteins were located across the TM domains 3-5; these domains are involved in ligand binding in other 7-transmembrane proteins (Kobilka 1992). The prediction that TM domains would be involved in ligand interaction was subsequently suggested in studies of larger OR repertoires, using bioinformatic approaches (Singer et al. 1996, Krautwurst et al. 1998, Zhao et al. 1998). In particular, bioinformatic studies identified a number of amino-acid residues which may be directly involved in ligand binding (Lapidot et al. 2001, Man et al. 2004), some of which have been confirmed experimentally by *in vitro* site-directed mutagenesis (Katada et al. 2005). All these studies suggest that TM domains 3-6 of OR proteins represent the binding pocket of ORs, and that variability across these domains reflects the diversity of odorant ligands that can be recognised.

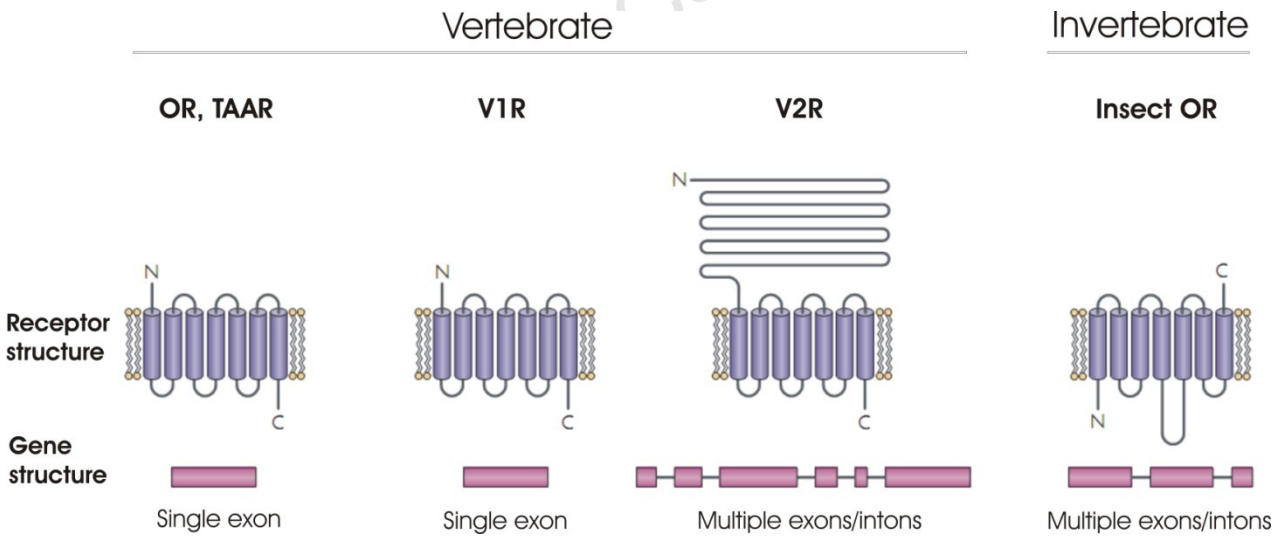
When suitable odorant ligands bind to ORs, a conformational change occurs and the olfactory-specific G-protein, $G_{\alpha olf}$ is stimulated, which in turn activates a cAMP-mediated signalling pathway in olfactory sensory neurons (Kato et al. 2008). The conformational changes that occur when an OR is activated are not fully understood; however, distinct residues have been identified in an intracellular loop and the C-terminal domain of ORs, both of which are necessary for functional signal transduction (Kato et al. 2008). Experimental evidence suggests that single ORs can respond to multiple chemical compounds, although with distinct binding affinities and specificities (Gaillard et al. 2002, Mombaerts 2004, Malnic 2007, Saito et al. 2009). Equally, individual odorants are found to elicit a response from multiple ORs (Malnic et al. 1999), and similar sets of odorants are thought to be recognised by ORs that share functional sequence motifs (Malnic et al. 2004).

University of Cape Town

Gene structure and genomic organisation

Like some other GPCR genes, OR genes have an intron-less coding region of approximately 1000 bps (Figure 1.4, Young & Trask 2002, Mombaerts 2004) and are classified into classes, families and sometimes subfamilies based on their sequence similarity (Glusman et al. 2000a, Zhang and Firestein 2002). OR genes form very compact units in the genome, with both the transcription start sites and the polyadenylation signals being usually short and situated in close proximity to the coding region (1-10 kb) (Qasba and Reed 1998, Vassalli et al. 2002, Fleischer et al. 2009).

Figure 1.4 Vertebrate and invertebrates olfactory receptor and gene structures (partly reproduced from Nei et al. 2008). Transmembrane structures of the main receptor families involved in olfaction. The N-terminals are extracellular in all receptor types, except insect ORs. OR, olfactory (or odorant) receptors; TAAR, trace-amine associated receptor; V1R and V2R, vomeronasal receptors type 1 and 2, respectively; Insect OR, insect olfactory receptor. Insect ORs have a 7TM structure similar to that of vertebrate ORs, but the receptor topology is inverted such that the N-terminal domain lies in the intracellular region (Benton et al. 2006); despite a number of similar anatomical features, insect and vertebrate ORs share no sequence similarity (Bargmann 2006). Insect ORs function as heretodimers, being co-expressed with the ubiquitously expressed odorant receptor Or83b (Larsson et al. 2004).



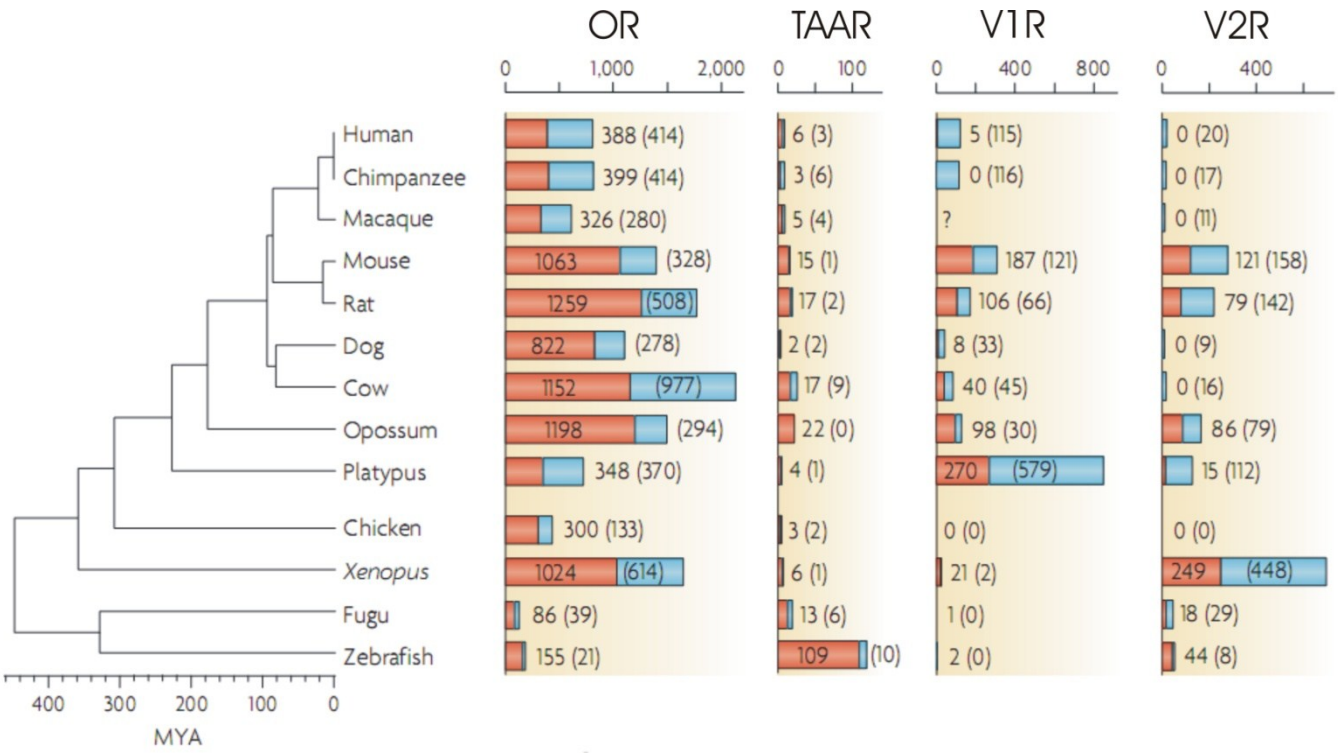
In the genomes currently characterised, ORs are organised in clusters, scattered on almost all chromosomes in the mouse and human genomes (Sullivan et al. 1996, Trask et al. 1998a, Rouquier et al. 1998, Glusman et al. 2001). Within clusters, the numbers of OR genes vary, and non-OR genes are not usually present. Inter-genic distances in OR clusters are variable - from less than 5kb to more than 50 kb - and a large proportion of interspersed repetitive elements is present (Xie et al. 2000, Glusman et al. 1996, Glusman et al. 2000b). These

repeats are thought to play a fundamental role in the evolution of OR clusters: repetitive elements play a causative role in gene duplication, favouring the transfer of genes to remote genomic locations, and even becoming an integrative part of the coding exon of ORs (Sosinsky et al. 2000). Indeed, current theories propose that the OR family evolves rapidly via a "birth-and-death" model where new OR genes arise through duplication and then differentiate in function in response to selection, while others lose function and undergo pseudogenization, or are deleted from the genome (Nei et al.1997, Nei & Rooney 2005, Nimura & Nei 2007). As a result, the size of any OR repertoire depends on diverse evolutionary forces together with the extent of duplication and inactivation events that characterise the evolution of a species' genome (Niimura & Nei 2007).

When Buck and Axel identified OR genes in rat, the size of the OR repertoire was estimated to be at least several hundred genes (1991). In the past 20 years, whole-genome sequencing projects have allowed for a comprehensive study of OR genes in several model species, and revealed unexpectedly large OR repertoires. Figure 1.5 shows the numbers of functional chemosensory receptor genes and pseudogenes in various vertebrate species (Nei et al. 2008). With ~1200 functional genes in rat, the OR multigene family is the largest in the vertebrate genome. Tetrapods have OR repertoires of circa 600-1500 genes, whereas fish have on average 100 genes (Niimura and Nei 2005b). The expansion of OR gene repertoires in tetrapods is thought to reflect the shift from aquatic to terrestrial environments in the Middle Devonian, some 395 MYA (Glusman et al. 2001).

University of Cape Town

Figure 1.5 Olfactory receptor repertoires across species (partly reproduced from Nei et al. 2008). The numbers of functional receptor genes (red bars) and pseudogenes (blue bars) are indicated for each receptor family, next to the respective bars; numbers of pseudogenes are in brackets. Data for the various receptor families were extrapolated from Nei et al.'s study (2008).



University of

Gene expression patterns

OR genes are mainly expressed in sensory neurons of the MOE, in a 'monogenic' way i.e. only a single receptor is expressed in each olfactory SN (Malnic et al. 1999, Touhara et al. 1999, Kajiya et al. 2001). In mice, OR expression is found to be monoallelic, with only the paternal or maternal allele of a particular OR gene being expressed in a given SN (Chess et al. 1994, Strotmann et al. 2000, Ishii et al. 2001). The mechanisms by which a SN selects a single OR, and represses the expression of all remaining ORs in the genome, is a matter of fascinating and controversial debate in the literature (Serizawa et al. 2003, Lomvardas et al. 2006, Fuss et al. 2007, Nishizumi et al. 2007). A given OR gene is usually expressed in a few thousand SNs within a particular region of the MOE (Ressler et al. 1993, Vassar et al. 1993, Iwema et al. 2004, Miyamichi et al. 2005), although a different expression pattern has been shown for a few OR genes (Strotmann et al. 1992, Pyrski et al. 2001). A small fraction of ORs is expressed simultaneously in the MOE and the VNO (Levai et al. 2006), as well as in the MOE and the SO (Kaluza et al. 2004).

Interestingly, a portion of OR genes is expressed in non-chemosensory organs (Feldmesser et al. 2006), a phenomenon referred to as 'ectopic expression', where the term 'ectopic' refers to 'a biological event or process that occurs in an atypical location or position within the body' (Feldmesser et al. 2006). Ectopically expressed ORs are found notably in sperm cells (Parmentier et al. 1992, Branscomb et al. 2000, Spehr et al. 2003, Fukuda and Touhara 2006), autonomic ganglia (Weber et al. 2002) or in the cerebral cortex (Otaki et al. 2003). Assuming that OR expression in the MOE is ancestral, evidence for ectopic OR expression raises the possibility that OR genes may perform additional functions in non-olfactory tissues (De la Cruz et al. 2009). However, with the exception of ORs that are expressed on the midpiece of sperm cells, and which appear to be involved in sperm chemotaxis (Spehr et al. 2003, 2006a), the non-olfactory roles of ORs are largely elusive.

Receptor families that complement the role of ORs in olfactory detection

The role of ORs in olfactory detection is complemented by the function of two additional GPCR families: vomeronasal receptors, VRs, which comprise two distinct subfamilies, V1Rs and V2Rs (Dulac and Axel 1995, Berghard and Buck 1996, Herrada and Dulac 1997, Matsunami and Buck 1997, Ryba and Tirindelli 1997, Tirindelli et al. 1998), and trace-amine associated receptors TAARs (Borowsky et al. 2001, Lindemann et al. 2005).

V1Rs

Like ORs, V1Rs belong to class A GPCRs, even though they lack significant sequence homology to rhodopsin-like receptors (Kristiansen 2004). In all species studied to date, V1Rs are extremely polymorphic but, unlike ORs, do not display specific sequence-motifs that could be described as diagnostic (Rodriguez et al. 2002). Given the structural similarities to ORs, the binding site of V1Rs is thought to lie within the TM regions of the receptor (Kristiansen 2004). V1Rs respond to a small number of receptor ligands, sometimes even to single compounds (Leinders-Zufall et al. 2000). Thus, in comparison to ORs, which respond to multiple odorants (e.g. Gaillard et al. 2002, Mombaerts 2004, Malnic 2007, Saito et al. 2009), the ligand spectrum of V1Rs appears limited. This has led to the hypothesis that the binding pocket of V1Rs may be more rigid and specific than that of ORs, which can accommodate numerous ligands.

In terms of gene structure, V1Rs - like ORs - are encoded by single-exon genes that are approximately 900 bps long (Figure 1.4, Zhang et al. 2004, Shi et al. 2005). The genomic

University of Cape Town

organisation of V1Rs has been studied in detail in rodents (Rodriguez et al. 2002, Zhang et al. 2004, 2007a), where V1Rs are organised in clusters across different chromosomes.

V1Rs are expressed in the apical portion of the VNO (Figure 1.2, Dulac and Axel 1995) where sensory neurons in the VNO appear to express a single V1R and in a monoallelic way (Rodriguez et al. 1999), as is the case with ORs. Interestingly, in humans, who lack a functional VNO, V1Rs are expressed in the MOE (Rodriguez et al. 2000), but their function outside the VNO is unclear.

V2Rs

V2Rs, on the other hand, belong to class C GPCRs, which are characterised by a long extracellular N-terminal domain (Pin et al. 2003). Like other class C GPCR receptors, the long N-terminal region of V2Rs is thought to represent the binding domain: a 'Venus flytrap-like' mechanism of ligand binding has been proposed for these receptors (Bridges and Lindsley 2008). Candidate ligands for V2Rs are thought to be non-volatile pheromones, such as peptides or proteins (Touhara 2007), and include major urinary proteins (MUPs) (Krieger et al. 1999, Chamero et al. 2007), major histocompatibility complex (MHC) peptides (Leinders-Zufall et al. 2004, He et al. 2008) and the exocrine gland-secreting peptide (ESP) family (Kimoto et al. 2005, 2007).

V2Rs have a complex intron/exon structure (Figure 1.4) which increases the length of individual genes (~20 kb) and complicates the identification of V2R sequences in genomic studies (Herrada and Dulac 1997, Matsunami and Buck 1997, Ryba and Tirindelli 1997, Yang et al. 2005a). Therefore, the current knowledge on V2R genomic organisation and evolution is still limited, even in model organisms.

V2Rs are expressed in the basal portion of the VNO (Figure 1.2, Herrada and Dulac 1997, Matsunami and Buck 1997, Ryba and Tirindelli, 1997) where they are often co-expressed with a particular subfamily of V2Rs, called *V2R2s*, whose members are present in high numbers in the basal VNO (Martini et al. 2001, Yang et al. 2005a, Silvotti et al. 2007). Interestingly, studies in mice suggest that V2Rs may form complexes with immune-system related proteins (Ishii et al. 2003, Loconto et al. 2003), although this model may not be generally applicable to all vertebrate V2Rs (Ishii and Mombaerts 2008). Recently a particular V2R subtype (V2r83) was found to be expressed outside the VNO, in the Grueneberg

University of Cape Town

Ganglion (GG) (Fleischer et al. 2006). The function of V2Rs expressed outside the VNO is yet to be established.

In general, the numbers of VR genes vary extensively among vertebrates (Figure 1.5), and particularly among mammals (Young et al. 2005, Grus et al. 2005, 2007, Young and Trask 2007). Functional V1Rs range from 0 (chimpanzee) to 270 (platypus), while functional V2Rs vary from 0 (human, chimpanzee, macaque, dog, cow) to 121 (mouse, Nei et al. 2008). To date, there is no clear relationship between the numbers of V1R and V2R genes, nor between the numbers of VR and OR genes.

VR genes would represent interesting targets for the investigation in this thesis, however two limiting factors of VRs dictated the choice to investigate OR genes. Firstly, it appears that at least one bathyergid species, the naked mole-rat *Heterocephalus glaber*, has a degenerate i.e. non-functional VNO (Smith et al. 2007). If this were a common feature across Bathyergidae species, VR genes may be not be expressed throughout the taxon (although 'ectopic' VR expression cannot be excluded, as is the case with the examples mentioned above - Rodriguez et al. 2000, Fleischer et al. 2006). Secondly, the genetic structure of VRs is sub-optimal for PCR-based characterisation: V1Rs do not display conserved sequence-motifs that could be used to confidently assign sequence identity (Rodriguez et al. 2002), while V2Rs have a complex intron/exon structure which would render characterisation arduous (Herrada and Dulac 1997, Matsunami and Buck 1997, Ryba and Tirindelli 1997, Yang et al. 2005a). In this regard, ORs represent ideal gene candidates because of their simple and distinctive genetic structure.

University of Cape Town

TAARs

Like ORs, trace-amine associated receptors, TAARs, are expressed in SNs of the MOE in a monogenic way (Liberles and Buck 2006); in addition to the MOE, a small number of TAARs are expressed in the GG (Fleischer et al. 2007). The coding region of TAARs is intron-less and of similar length to that of ORs and V1Rs (Figure 1.4, Lindemann et al. 2005). When compared to other chemosensory receptor families, TAARs are less numerous e.g. six functional genes in human and 15 in mouse (Figure 1.4, Nei et al. 2008). Mouse TAARs are activated by certain amine ligands which are present in urine in gender- and stress-dependant concentrations, suggesting a possible role for TAARs in detecting pheromones that carry social cues (Liberles and Buck 2006).

1.2 The African mole-rats, Bathyergidae

The etymology of the word Bathyergidae comes from the Greek *bathys*, meaning deep, and *ergo*, meaning to work. The Bathyergidae are a family of subterranean rodents endemic to sub-Saharan Africa, where they occur across a wide range of habitats and soil types (Bennett & Faulkes 2000). Interest in the family peaked after Jarvis (1981) described the social system of one of its species - *Heterocephalus glaber*, commonly known as the naked mole-rat (Figure 1.6) – as being 'eusocial' (Batra 1966, Michener 1969, Wilson 1971). Over the past decade, interdisciplinary research has extended to include all the genera within the Bathyergidae, revealing a number of unique and fascinating aspects of African mole-rat evolutionary biology including their behaviour, ecology, neurobiology, systematics and physiology. The Bathyergidae are unique amongst mammals because they display the complete range of social systems, from strictly solitary, to social and eusocial species (Bennent and Faulkes 2000). A synthesis of the salient aspects of African mole-rat phylogeny, ecology, sociality and physiology is reported in this section, with a particular emphasis on the role of olfaction in the taxon.

University of Cape Town

Figure 1.6 Examples of the six genera of African mole-rats Species names are indicated for each image; the photographer is referenced in brackets when known.

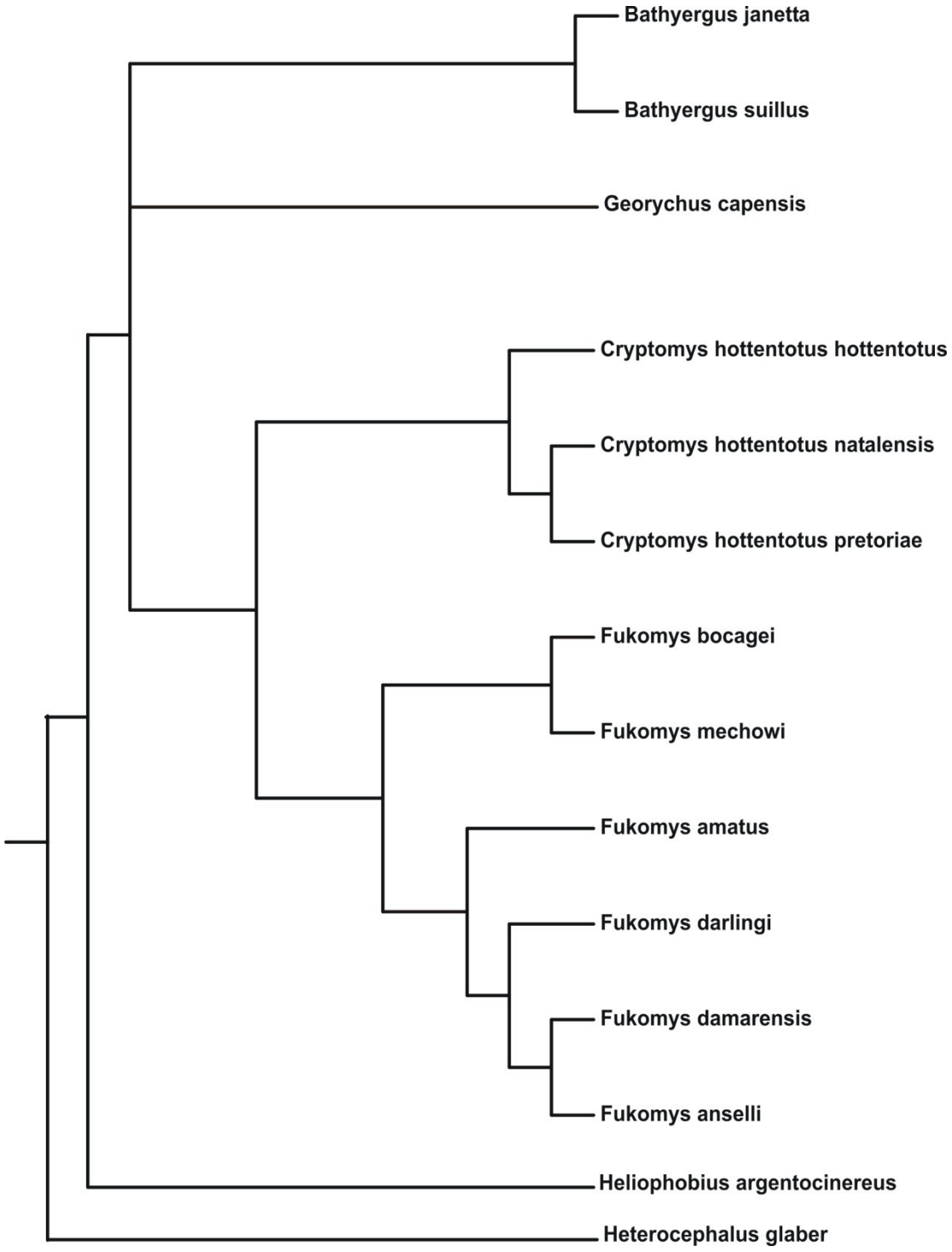
Phylogenetic relationships in the Bathyergidae

The order Rodentia is divided into five sub-orders: the Myomorpha, the Sciuromorpha, the Castorimorpha, the Anomaluromorpha and the Hystricomorpha or Hystricognathi (Wilson and Reeders 2005). Based on morphometric features such as the arrangement of jaw muscles and the shape of the skull (Wood 1985), bathyergids appear to be closest to hystricognath rodents (Honeycutt et al. 1991) and some of the characteristics of their reproductive cycle further support the placement of Bathyergidae within Hystricomorpha (e.g. Faulkes et al. 1990). Numerous studies consistently support a monophyletic origin for the Bathyergidae, which is estimated to have diverged from its common ancestor as long as 49 million years ago during the Eocene (Nedbal et al. 1994, Huchon and Douzery 2001, Blanga-Kanfi et al. 2009).

The Bathyergidae is further divided into two sub-families, Bathyerginae and Georychinae, which traditionally included five genera, based on dental parameters (Roberts 1951, De Graaf 1981). Bathyerginae contained a single genus, *Bathyergus*, whilst Georychinae traditionally contained the four genera *Heterocephalus*, *Heliophobius*, *Georychus* and *Cryptomys*. Recently, a new bathyergid genus - *Fukomys* - was described, containing several species that were formerly placed in the genus *Cryptomys* (Kock et al. 2006). *Fukomys* is estimated to have diverged from *Cryptomys* ~10-11 MYA (Ingram et al. 2004) and currently represents the most speciose genus within the family (Ingram et al. 2004, Van Daele et al. 2007). Relationships between Bathyergid genera and species analysed in this study are summarised in Figure 1.7.

University of Cape Town

Figure 1.7 Phylogenetic relationship of African mole-rat species analysed in this study The tree represents a consensus between trees obtained using different models, based on 12s RNA and cyt-b data, which all share the same overall topology (Faulkes et al. 1997, Faulkes et al. 2004, Ingram et al. 2004, Van Daele et al. 2007). Trees constructed with 12S rRNA data (Ingram et al. 2004) and rooted with the African cane rat *Thryonomys swinderianus*, recover a *Bathyergus* + *Georychus* clade (not shown here), albeit with weak support (Ingram et al. 2004).



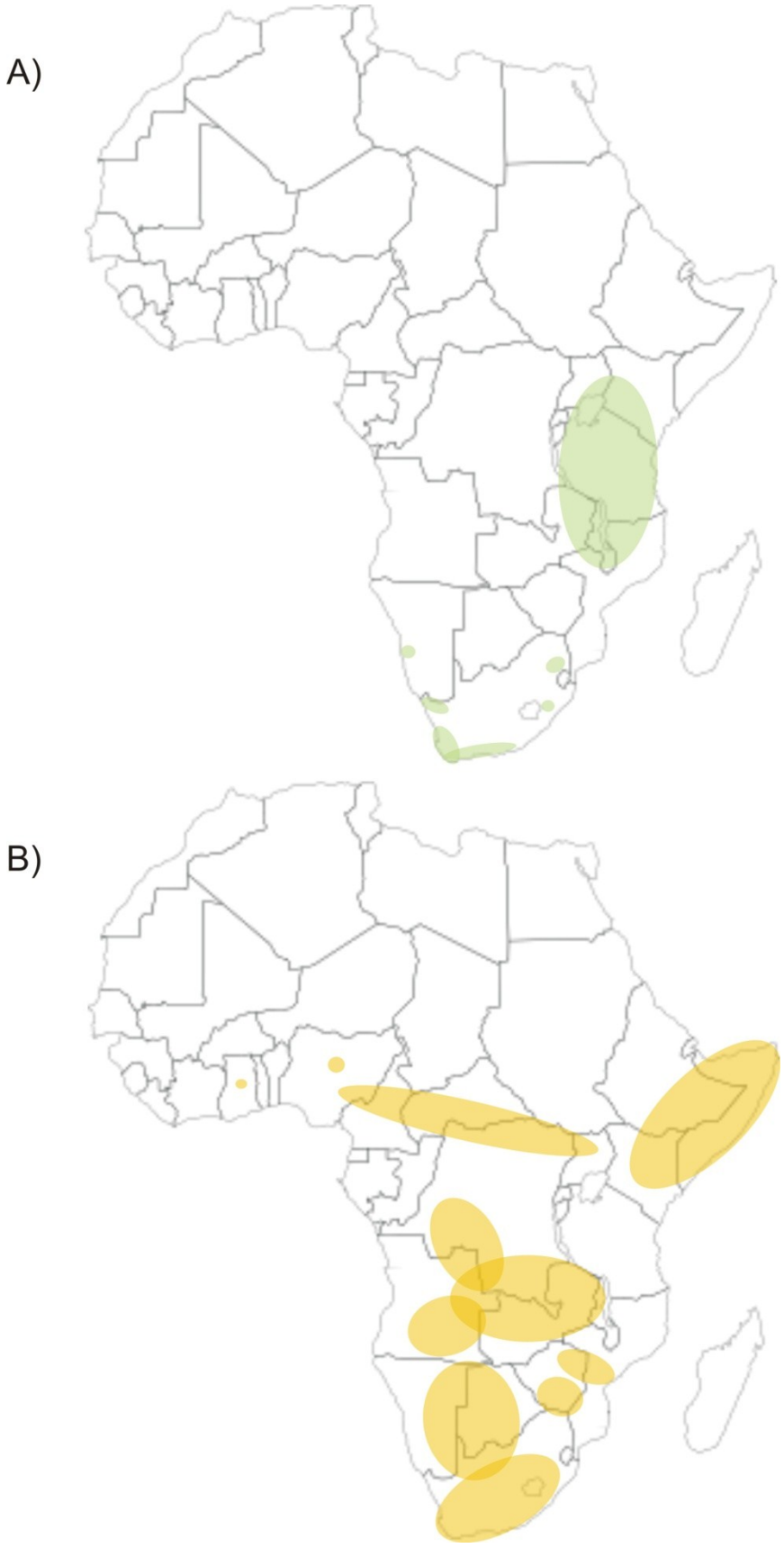
Ecology and the evolution of sociality in the Bathyergidae

Across sub-Saharan Africa, African mole-rats occupy a wide range of habitats, from mesic to xeric, and soil types, from coarse arenosols (sands) to fine clays (Bennett and Faulkes 2000). In many areas species are distributed across a number of ecological gradients including rainfall patterns, altitudinal and vegetation types, with the presence of geophytic plant species being the only common denominator for all Bathyergidae. Geophytes are plants that possess underground storage organs – such as corms, tubers, bulbs or rhizomes – and constitute the main source of nutrition and water for all African mole-rats; only a few species in the genera *Georychus* and *Bathyergus* complement their diet with aerial vegetation (Davies and Jarvis 1986, Bennett 1988).

Within the genera *Bathyergus*, *Georychus* and *Heliophobius*, species occupy mesic habitats where precipitation is relatively high (>400 mm per annum) (Figure 1.8 A). These species are all solitary, with multiple-occupancy of the same burrow system restricted to the breeding period, and individuals are generally larger in size than their social counterparts (Bennett & Faulkes 2000). In contrast, species in the genera *Cryptomys, Fukomys* and *Heterocephalus* are highly social and occur in both mesic and xeric regions distributed from the Horn of Africa, central and west Africa, through south-central Africa down to the southernmost tip of Africa (Figure 1.8 B) (Bennett and Faulkes 2000). Social mole-rats tend to be more successful than the solitary ones in occupying wider distributional ranges, to include areas where rainfall may be scarce and unpredictable (<200 mm per annum for some species) (Jarvis and Bennett 1990, 1991). In social species, colonies are characterized by the presence of a single reproductive female at any one time, together with unambiguous reproductive division of labour along a linear hierarchy amongst the other colony members, with non-breeding individuals cooperating in the raising of litters. The size of social mole-rat colonies can be very large, with up to 41 individuals reported for the genus *Cryptomys* (Jarvis and Bennett 1993), and about 90 individuals per colony being common for *Heterocephalus* (Bennett and Faulkes 2000).

University of Cape Town

Figure 1.8 Bathyergidae distribution map A) Distribution of the solitary genera *Bathyergus*, *Georychus* and *Heliophobius* - green; B) distribution of the social genera *Cryptomys*, *Fukomys* and *Heterocephalus* – yellow (Bennett and Faulkes 2000).



African mole-rats are a unique mammalian family in that they display the complete range of species sociality, from the strictly solitary genera *Bathyergus*, *Georychus* and *Heliophobius*, to the cooperative breeding ones *Cryptomys, Fukomys* and *Heterocephalus*. Some of the species in the latter two genera have been described as "eusocial", but this definition has been the subject of much debate (Jarvis and Bennett 1993, Burda et al. 2000, Bennett & Faulkes 2000, Burland et al. 2002).

Eusociality was first described by entomologists as a social colonial system fulfilling three conditions: 1) a reproductive division of labour with only a few breeding individuals; 2) overlap between generations of non-breeding helpers and 3) cooperative rearing of offspring and maintenance of the colony (Batra 1966, Michener 1969, Wilson 1971). Around the mid-'90s, the term 'eusocial' was the subject of an intensive 'semantic' debate (Crespi and Yanega 1995, Keller and Perrin 1995, Sherman et al. 1995, Reeve et al. 1996, Wcislo 1997), which later became a conceptual discussion of the fundamental characteristics of cooperative breeding animal societies (Costa and Fitzgerald 2005, Lacey and Sherman 2005). In this theoretical debate, three dominant themes for characterising 'eusociality' were considered: i) those that tended to differentiate eusociality from other social systems (e.g. Crespi and Yanega 1995), ii) those that considered eusociality as part of a continuum of cooperative social structures (e.g. Sherman et al. 1995) and iii) those that relied on phylogenetic relationships to assign social structure (e.g. Wcislo 1997). Depending on which definition of eusociality is applied to a species, the phenomenon of eusociality appears to be either very rare and exclusively entomogenic, or relatively common, with some unexpected mammalian species being considered putatively 'eusocial' e.g. humans (Foster and Ratnieks 2005), whales (McAuliffe and Whitehead 2005). In the specific context of Bathyergidae, Burda (1999, 2000) suggested the following additional criterion to be added to Batra's (1966) original definition of eusociality for describing eusocial bathyergids: permanent phylopatry. According to Burda's definition most, if not all, *Cryptomys* and *Fukomys* species should be considered eusocial, together with the naked mole-rat *H. glaber*. Other authors consider different criteria to describe eusocial mammals, such as colony size and the level of social cohesion within colonies e.g. Jarvis et al. (1994), Faulkes et al. (1997), Faulkes (1998), Wallace and Bennett (1998). Following these authors, only two species of African mole-rat can unambiguously be called eusocial, namely the

University of Cape Town

Damaraland mole-rat *Fukomys damarensis* and the naked mole-rat *Heterocephalus glaber*. Both sides of this debate (Burda 2000, Bennett and Faulkes 2000) do, however, agree that direct and undisputed field evidence for eusociality in African mole-rats is currently available only for *H. glaber* (Jarvis 1981, Brett 1991, Braude and Ciszek 1998) and *F. damarensis* thus far (Jarvis and Bennett 1993, Jarvis et al. 1994). Therefore, in this thesis only these two species are considered 'eusocial', whilst the other species of the genera *Cryptomys* and *Fukomys* are classified as 'social'.

Among the majority of authors studying the evolution of sociality in the Bathyergidae, there is general consensus that sociality is a derived trait with eusociality being its endpoint (e.g. Jarvis and Bennett 1990, 1991, Faulkes et al. 1997). Notwithstanding the basal position and evolutionary age of the eusocial naked mole-rat lineage (Figure 1.7), this view is motivated by two main factors: i) the assumption that a solitary lifestyle is plesiomorphic (ancestral) amongst extant subterranean rodent families because the majority of them are solitary (Nevo 1979) and ii) the fact that the earliest bathyergid fossils found were large in size (Lavocat 1974), and large body size is characteristic of most extant solitary mole-rat species (Jarvis and Bennett 1990). A scenario where sociality represents a derived trait supports one of the most accepted hypotheses on the evolution of eusociality in African mole-rats, the Aridity-Food Distribution Hypothesis, AFDH (Jarvis and Bennett 1991, Jarvis et al. 1994, Faulkes et al. 1997, Faulkes 1998). According to the AFDH, (eu)sociality evolved in response to the ecological constraints of arid habitats, where (i) the costs of burrowing in incredibly hard soils is high for most of the year, and (ii) food resources are patchy and widely spaced. In such habitats, group living, cooperative foraging, reproductive altruism and communal care are highly adaptive and act as preconditions for the evolution of (eu)sociality. Within-species studies on variation in social cohesion, colony structure and parentage in the common mole-rat *Cryptomys hottentotus hottentotus* - a social species whose distribution follows a strong mesic to arid environmental cline - provide further support for the AFDH (Spinks et al. 2000, Bishop et al. 2004).

An alternative scenario on the origin of bathyergid (eu)sociality is argued by Burda (1997, 1999, 2000), where sociality is the ancestral condition. Burda's view is motivated by a number of different ideas, including Kleiman's findings (1974) that

University of Cape Town

most hystricomorph rodents are social e.g. the Southern mountain cavy *Microcavia australis*, the Chacoan mara *Dolichotis salinicola*, the degu *Octodon degus* (see also Ebensperger and Blumstein 2006), and that the evolution of sociality is thought to be independent of ecological factors. In addition, Burda argues that social living, either in groups or in pairs, would have dramatically increased the chances of survival during involuntary migrations, transoceanic rafting and founding of viable populations from the African to the South American continent. Importantly, all these aspects are potentially associated with the successful radiation of hystricognath lineages in South America (Burda 1997, 1999, 2000). Thus, in Burda's view, cooperative 'monogamy' is more likely to represent an ancestral trait amongst bathyergids; a vital pre-adaptation for the successful transition to living underground. Due to constraints on individual dispersal the subterranean lifestyle reinforced this condition, leading to the rapid accumulation of overlapping generations that characterises all extant social mole-rats (Burda 2000).

African mole-rat sensory physiology and olfaction

African mole-rats are highly adapted to their subterranean existence, and above ground activity is generally rare and only reported for a few solitary species (Jarvis and Bennett 1991). Burrows represent a safe and thermostable niche, but they are also characterized by a rather hostile set of physical conditions such as darkness, high relative humidity, high carbon dioxide levels and low oxygen concentrations (Kennerly 1964, Darden 1972). Furthermore, the subterranean lifestyle imposes important constraints on foraging and the location of mates. This extreme set of conditions has led to the evolution of a number of morphological, physiological and behavioural adaptations common to all mole-rat species.

In the absence of visual cues, communication among individuals in a subterranean environment requires specific sensory adaptations (Eloff 1958, Bennett & Faulkes 2000). Auditory sensitivity is limited to low frequency sounds in mole-rats e.g. despite a broad range of vocalisations being described for the species, naked mole-rats appear to be have limited sensitivity to a range of sounds when compared with their surface-dwelling rodent counterparts (Heffner and Heffner 1993, Pepper et al. 1991). Tactile cues within the burrow system e.g. for orientation and involving individual contact, are conveyed through highly sensitive vibrissae that are found in

University of Cape Town

high concentrations around the head and along the body and tail in all Bathyergids (Thigpen 1940, Crish et al. 2003). With regards to their sense of smell, experimental evidence reveals the presence of a well-developed olfaction which has emerged as fundamental to the evolutionary success of African mole-rats (Faulkes 1990, Judd and Sherman 1996, Heth et al. 2002a, 2002b, 2004, Lange et al. 2005). For example, behavioural studies conducted on three different families of subterranean rodents - the East-Mediterranean blind mole-rats *Spalax* (Spalacidae), the African mole-rat genera *Fukomys* and *Heterocephalus* (Bathyergidae), and the South American Coruro *Spalacopus* (Octodontidae) - revealed that all species from all three families were able to detect plant exudates, released by roots of edible plants into moist soils (Heth et al. 2002a, Lange et al. 2005). Individuals were able to detect odorant substances produced by edible plants over a distance of at least 30 cm, and to preferentially dig in their direction (Lange et al. 2005), suggesting that subterranean rodents may intentionally orientate their digging towards food sources to minimise the energy investment necessary for successful foraging. Interestingly, this role for olfaction in foraging seems to have evolved via convergence in subterranean rodents, since it is found across three different rodent families and continents (Heth et al. 2002a). Importantly, these findings are not consistent with a number of studies suggesting that African mole-rats generally forage randomly (e.g. Lovegrove and Wissel 1998, Brett 1991, Jarvis et al. 1998, Spinks et al. 2000).

Studies on the naked mole-rat have revealed the use of olfactory cues in recruiting colony members to food sources, by laying down odour trails (Judd and Sherman 1996). Furthermore, naked mole-rats make significant use of olfactory cues during colony member and kin recognition interactions (Faulkes 1990, O'Riain and Jarvis 1997, Reeve and Sherman 1991, Jarvis 1991); and complex scent marking rituals are used in common nesting and latrine areas within their extensive burrow systems (Jarvis and Sherman 2002). Additionally, mole-rats of the genus *Fukomys spp.* are able to discriminate between kinspecific and heterospecific odours (Heth et al. 2002b, 2004) via a proposed 'self-referent matching' mechanism (Holmes and Sherman 1983), and use this information to both reinforce individual and group recognition rituals and to limit incestuous matings (Burda 1995).

University of Cape Town

In the last decade, numerous studies have explored neurobiological and molecular aspects of mole-rat sensory physiology and behaviour, concentrating on the naked mole-rat as a new mammalian model system. For example, skin and hair innervations have been thoroughly characterised in the species, with regards to their function in touch-guided orienting behaviour (Cris et al. 2003), but also in connection with thermoregulation and pain tolerance (Park et al. 2003). Indeed, naked mole-rats appear unable to regulate their body temperature - a typically 'poikilothermic' trait (Buffenstein and Yahav 1991a, 1991b) - and appear quite insensitive to a number of standard pain stimuli, presumably due to the observed lack of specific neuropeptides that are usually associated with thermoregulation and pain in other mammals (Park et al. 2003). Other recent studies have focused on the distribution of various hormones and hormone receptors in the brain e.g. vasopressin, androgen and oxytocin receptors (Rosen et al. 2007, Holmes et al. 2008, Kalamatianos et al. 2010). Arginine vasopressin (VP) is associated with social behaviors, including pair bonding, parental behavior, and dominance-subordinance in many species (Goodson and Bass, 2001, Lim et al. 2004, Donaldson and Young 2008). In naked mole-rats, patterns of vasopressin receptor expression reveal that reproductive individuals have an additional area of VP expression when compared to subordinates (Rosen et al. 2007). Surprisingly, the opposite scenario characterises the distribution of androgen receptors, with fewer receptors present in breeders than in subordinates (Holmes et al. 2008), suggesting that reproductive individuals may have a reduced response to androgens e.g. testosterone. Given the established relationship between testosterone and aggression (Wingfield 2005), it is possible that androgen receptors are decreased in breeding naked mole-rats to facilitate life in a eusocial society, in which many animals live in close quarters with remarkably few agonistic encounters (Clarke and Faulkes 1997). A recent study on oxytocin and oxytocin receptor (OTR) binding sites in the naked mole-rat and the solitary Cape mole-rat *Georychus capensis* revealed an increase in production of oxytocin and its binding sites in the social naked mole-rat (Kalamatianos et al. 2010). Oxytocin is fundamental to pro-social behaviour in mammals (Insel and Young 2000, Campbell 2008). An increased role for oxytocin in the naked mole-rat is consistent with the extreme social behaviour of the species and the ability to form reproductive bonds, while the opposite molecular pattern observed in the solitary Cape mole-rats is in line with their rare pro-social behaviour (Kalamatianos et al. 2010).

University of Cape Town

1.3 Research objectives and thesis structure

In the context of the contemporary molecular studies on bathyergid physiology and behaviour, and given the significant role of olfaction in the ecology and behaviour of the Bathyergidae, the main aims of this study are to:

i) Explore the genetic basis of olfaction by isolating and characterising a subset of OR genes from 14 extant species of the Bathyergidae
ii) Contextualise bathyergid OR genes in a phylogenetic framework based on the available information on mammalian OR subgenomes
iii) Investigate the evolutionary mechanisms that govern OR gene evolution in the Bathyergidae
iv) Identify the relative contributions of environmental and social traits in driving the adaptive evolution of the bathyergid OR gene repertoire

This thesis is structured into five main chapters, including the General Introduction (Chapter 1). Using a PCR-based and bioinformatic approach, Chapter 2 addresses points i) and ii) by describing the isolation and characterization of OR genes in 14 extant species of African mole-rat. To contextualise and classify the first subset of OR genes to be isolated in a subterranean mammalian family, OR phylogenies are inferred both across mole-rat species, and across mammalian OR subgenomes. In Chapter 3, point iii) is explored by determining the selective forces that have shaped the bathyergid OR repertoire isolated in this study. Results are interpreted in the framework of Nei's 'birth and death' model of evolution (Nei et al. 1997), which characterises the evolution of OR subgenomes. Finally, using a modern bioinformatic approach Chapter 4 explores the differential selective roles of bathyergid life history traits and environmental niche specialisation on OR gene evolution and diversification (point iv). A synthesis and concluding remarks on this study are presented in Chapter 5.

University of Cape Town

Chapter 2: Isolation and characterization of olfactory receptor genes, ORs, in African mole-rats

2.1 Introduction

In 1991, Linda Buck and Richard Axel first described Olfactory Receptor genes, a discovery for which they were awarded the Nobel Prize for Physiology or Medicine in 2004. In their pioneering study, Buck & Axel (1991) based their experimental approach on three assumptions. First, odorant receptors were predicted to belong to the 7 trans-membrane (7-TM) G-protein coupled receptor (GPCR) superfamily. Second, the large number of distinct odorant molecules was likely to reflect into a great variability of odorant receptors, which were therefore expected to be encoded by a multigene family. Third, expression of olfactory receptors should be limited to the olfactory epithelium.

Based on their intuition, Buck and Axel (1991) used degenerate PCR primers to target the conserved regions of 7-TM GPCR genes in cDNA prepared from rat olfactory epithelium, as described in chapter 1. In this way, Buck and Axel (1991) were able to isolate and characterise 18 novel genes which displayed known motifs of GPCRs, but which also shared some unusual features and were therefore representative of a new family of receptors - the Olfactory Receptor multigene family.

University of Cape Town

Techniques to characterize Olfactory Receptor genes

Since the seminal work of Buck and Axel (1991) a growing body of evidence has accumulated that corroborate the original findings (Mombaerts 1999a & 2004, Ache and Young 2005). The increasing availability of whole-genome sequence data made it possible to characterise entire OR repertoires in several model species, using *in silico* tools. Extensive data-mining searches like the ones described by Niimura and Nei (2003, 2005a, 2007) have led to the characterisation of entire OR subgenomes in many model organisms. These include several mammalian OR repertoires, such as human (Glusman et al. 2001, Zozulya et al. 2001, Niimura and Nei 2003), mouse (Zhang and Firestein 2002, Niimura and Nei 2005a, Zhang et al. 2007a), dog (Quignon et al. 2005), and other mammals (Niimura and Nei 2007, Kishida 2008), as

well as a number of non-mammalian chordate OR repertoires (Alioto and Ngai 2005, Niimura and Nei 2005b, Niimura 2009). These *in silico* studies led to an increased understanding of OR gene diversity and phylogeny. Nevertheless, it should be noted that these whole-genome approaches are limited to the study of model species, and are therefore not able to answer evolutionary questions in non-model organisms.

An alternative approach to study OR repertoires in non-model species relies on the original principles described by Buck and Axel (1991) using PCR-based methods. This approach has remained a powerful tool to investigate hypothesis-driven questions on the evolution of OR repertoires in poorly characterized genomes. For example, this approach has been used extensively in primates, revealing low OR gene diversity and diminished olfactory abilities in comparison with rodents (Rouquier et al. 2000), with a correlation of a deteriorated sense of smell with the acquisition of full trichromatic vision (Gilad et al. 2004, but see Matsui et al. 2010). Similarly, a PCR-based study on marine vertebrate ORs supports the hypothesis of a loss of olfactory abilities in fully marine-adapted mammals (cetaceans) in comparison with semi-adapted marine vertebrates (sea lions and sea turtles) (Kishida et al. 2007). In birds, on the other hand, OR diversity is related to unexpected olfactory sensitivity (Steiger et al. 2008), in particular among nocturnal bird species when compared with their diurnal relatives (Steiger et al. 2009). The link between OR repertoires and olfactory ability is discussed in further detail in Chapter 3.

In the past two decades, the advent of modern PCR machines and sequencing techniques has simplified the PCR-based experimental approach to gene characterisation. Notably, research has shown that PCR-based approaches for the characterisation of OR genes can yield similar results to those of whole-genome estimates. For example, Gilad et al. (2004) and Malnic et al. (2004) found similar proportions of OR genes: pseudogenes in humans using a PCR-based and an *in silico* approach, respectively. An accurately designed PCR-based experimental procedure will result in minimal numbers of polymerase or sequencing artefacts (Whinnet & Mundy, 2003). Additionally, information on ORs available in public databanks is used routinely to contextualise and validate the results obtained *in vitro*.

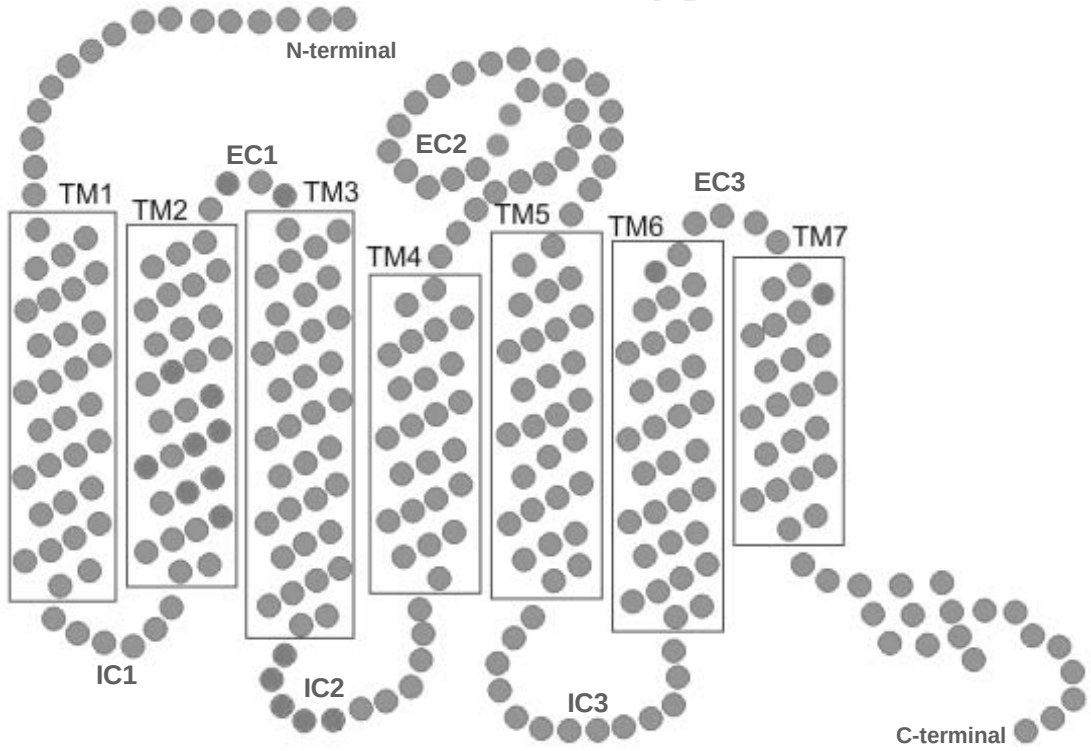
University of Cape Town

2.1.1 Olfactory Receptor genetic structure

Like other members of the GPCR family, OR genes encode seven hydrophobic amino-acid stretches, corresponding to the trans-membrane receptor domains, TM1-7 (Buck and Axel 1991, Mombaerts 1999b). The trans-membrane domains are intercalated by three intracellular and three extracellular loops (IC1-3 and EC1-3 respectively), which contain a number of amino acid residues common to all GPCRs. For example, the conserved cysteine residues in EC2 and TM3 form a structural disulfide bond in all GPCRs, (Katada et al. 2005 - Figure 2.4.b). The start of the OR gene product consists of an extracellular N-terminal, whilst a C-terminal domain lies at the other receptor's extremity (Figure 2.1).

University of Cape Town

Figure 2.1 Typical olfactory receptor structure (redrawn from Katada *et al.* 2005) Abbreviations stand for the following: TM trans-membrane domain, EC extra-cellular and IC intracellular domain.



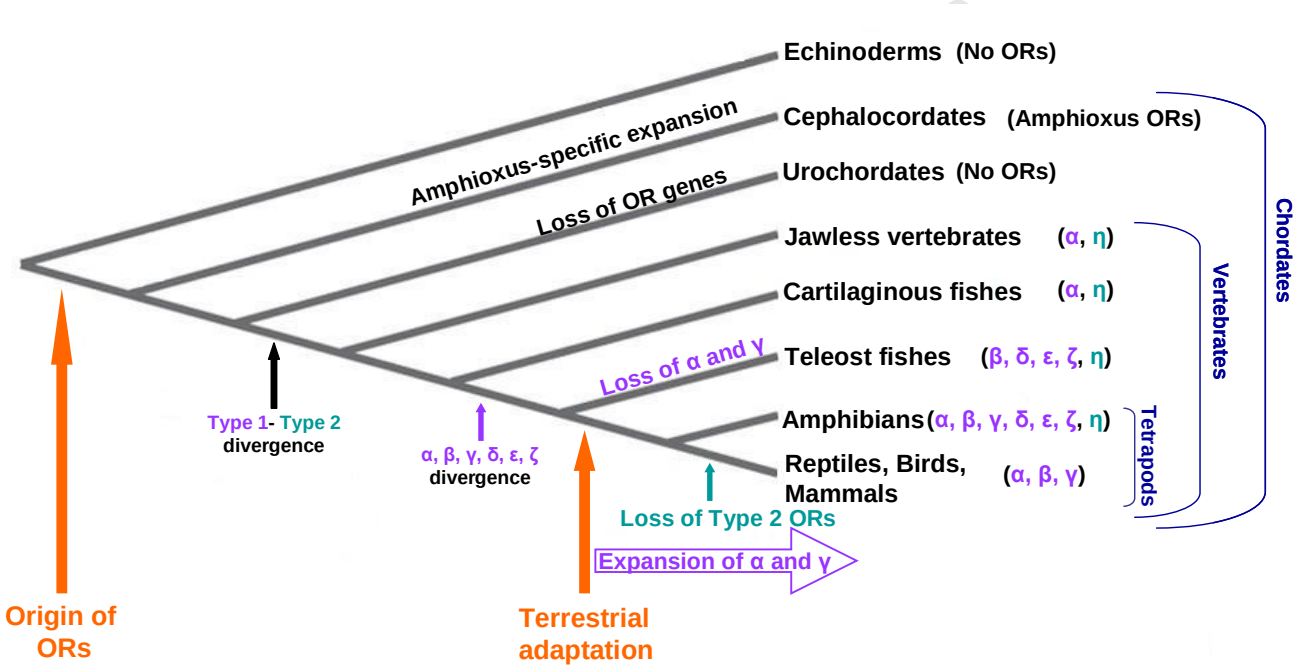
Unlike other GPCR genes, OR genes have a peculiar intron-less nature which makes them relatively short (~1000 bps, Young & Trask 2002, Mombaerts 2004), as compared to, for example, mouse opsin genes, which are ~9.5 kb long including introns and flanking sequences, (Al-Ubaidi et al. 1990). Non-coding exons situated up- and down-stream the coding sequence (i.e. the transcription start site and the polyadenylation signal, respectively), together with the corresponding introns, are usually short and located in close proximity to the coding region (1-10 kb) (Fleischer et al. 2009). The non-coding exons situated upstream the OR coding region can undergo alternative splicing, resulting in different mRNA isoforms which nonetheless translate into the same OR proteins (Volz et al. 2003, Young et al. 2003). Distinctive features of OR genes include typical amino-acid motifs, such as the MAYDRFVAIC, SY and KAFSTCASH motifs in TM3, TM5 and TM6, respectively (Appendix I.1, Zhang & Firesetin 2002, Godfrey et al. 2004, Zhang et al. 2007a); these motifs may vary slightly across species, but they represent established indicators of OR sequence identity in all species studied to date (Fleischer et al. 2009). A high concentration of sequence variability, spanning TM domains 3 to 6, is typical of OR genes. This region constitutes the ligand-binding domain of ORs (Katada et al. 2005), where high levels of polymorphism are thought to be proportional to the range of odorant molecules that bind to the receptors (Niimura and Nei 2006, Kishida 2008).

University of Cape Town

2.1.2 Evolution and classification of olfactory receptor genes in Vertebrates

A recent comprehensive survey of whole-genome data from 23 Chordate species subdivides OR genes in two groups: Type 1 and Type 2 (Niimura 2009). The divergence between these two types of OR genes is thought to have been in place in the most recent common ancestor of all Vertebrates (Niimura and Nei 2005b). Type 1 genes are partitioned in six subgroups, named α - ζ, whereas Type 2 genes comprise five orthologous groups, namely η, θ1, θ2, κ, and λ. Of the latter Type 2 genes, only η appears to be an authentic OR group, whereas θ1, θ2, κ, and λ fail to display typical OR features (e.g. duplicated genes) and their function is unknown (Niimura and Nei 2005b, Niimura 2009). The most parsimonious explanation on the evolution of OR genes and of the α - η olfactory subgroups is schematised in Figure 2.2.

Figure 2.2 Olfactory receptor origin and evolution (redrawn from Niimura 2009) Schematic diagram of the evolution of OR subgenomes in Chordates. Colours indicate Type 1 (**purple**) and Type 2 (**blue**) OR genes. Symbols in brackets (α - η) indicate the OR groups present in different taxonomic lineages. Olfactory Receptor genes, ORs, originated in the most recent common ancestor (MRCA) of all Chordates. In Amphioxus, ORs evolved in a lineage-specific manner: Amphioxus OR genes are highly divergent from vertebrate ORs and are characterized by long C-terminal tails. No OR genes are found in the Urochordate lineage. In Teleost fishes, ORs from groups α and γ were lost, and the number of OR genes in the other groups appears to be highly variable. In Tetrapods, on the other hand, groups α and γ have expanded dramatically, probably due to the importance of olfaction in terrestrial habitats.
It is suggested that α and γ ORs are for detecting volatile odorants, whilst δ, ε, ζ and η are thought to recognise water-soluble compounds, and group β genes have been proposed to bind to both hydrophilic and hydrophobic odorants (Niimura 2009, Niimura and Nei 2005b).



Mammalian ORs: Classes

Traditionally, mammalian OR genes are subdivided in Class I and Class II (Freitag et al. 1995, Glusman et al. 2000a), which both belong to Type 1 ORs (Class I corresponds to groups α and β, Class II to group γ; Niimura and Nei 2005b, 2006).

Class I receptors were first described in catfish (Ngai et al. 1993) and were initially regarded as 'fish-like' ORs, whereas Class II ORs were thought to be 'mammalian-like' (Freitag et al 1998). This classification was based on the original finding that the frog (*Xenopus laevis*), which has two different nasal cavities - one water-filled and one air-filled – had two different sets of ORs (Freitag 1995). The set of OR genes

expressed in the water-filled cavity displayed sequence homology to the known fish OR genes (Class I), whereas genes expressed in the air-filled cavity resembled the mammalian OR genes identified at that time (Class II) (Freitag 1995). As a result, Class I ORs were thought to recognise water-borne odorants, whereas Class II receptors were thought to detect volatile compounds (Zhang and Firestein 2002).

In 2000 Glusman et al. (2000a) identified a number of Class I and Class II genes in various vertebrate species, including a range of mammalian, marsupial, monotreme, amphibian, bird and fish ORs. Although the majority of mammalian ORs identified in vertebrates belong to Class II (Niimura 2009, Hayden et al. 2010), a substantial range of Class I ORs were characterised in Human (Glusman et al. 2001) and Mouse (Zhang and Firestein 2002) using genomic approaches. The subdivision of OR genes in Class I and Class II was therefore extended to all vertebrate ORs, even though fish do not possess Class II genes, and the function of Class I 'fish-like' genes in mammals remains enigmatic, since they were thought to recognise water-soluble odorants (Fleischer et al. 2009).

Phylogenetic analyses (Niimura and Nei 2006) revealed that OR genes in fish are only distantly related to tetrapod Class I ORs. Fish ORs belong to groups β, θ, ε and ζ, whereas mammals possess ORs from groups α and β. This suggests that only β Class I ORs are truly orthologous in fishes and mammals (Niimura 2009) (Figure 2.2). Indeed, genes from group α have been suggested to detect air-borne molecules (Niimura and Nei 2005b, Niimura 2009), whilst group β genes, are thought to recognise both hydrophilic and hydrophobic odorant molecules (e.g. alcohol), therefore being useful in both the aquatic and terrestrial environments (Niimura 2009). Class II genes (group γ), on the other hand, have expanded enormously in mammals, presumably due to the importance of olfaction in terrestrial life (Niimura and Nei 2005b, Niimura 2009).

Mammalian ORs: Families

Using genetic similarity criteria, the mammalian OR repertoire has been further subdivided into specific gene families (Glusman et al. 2000a, Zhang and Firestein 2002). In general, GPCRs that display similar sequences are known to share functional traits (Vassilatis et al. 2003), and the same is expected with ORs (Malnic et

University of Cape Town

al. 2004). As a general rule, OR genes that share ≥40% sequence similarity are regarded as belonging to the same OR family (Glusman et al. 2000a, Godfrey et al 2004). According to this criterion, the mammalian OR subgenome is conventionally partitioned into 17 families: four Class I - namely families 51, 52, 55 and 56 - and 13 Class II – named families 1 to 13 (Glusman et al. 2000a, Warren et al. 2008). Although the differential functions of these families and the range of odorants they can recognise is poorly understood (Nei et al. 2008), Zarzo (2007) suggested that each family might detect a particular class of odorant molecules. This hypothesis is based on the resemblance between the number of estimated dimensions in odour space (17-19 odour dimensions according to Jeltema & Southwick 1986, Abe et al. 1990) and the number of gene families in the OR subgenome (17 OR families, Glusman et al. 2000a, Niimura &Nei 2003).

2.1.3 Aims of this Chapter

This chapter details the isolation and characterization of OR genes, using a PCR-based approach, in 14 extant species of African mole-rats. The experimental procedure was carefully designed to minimise the generation of artefactual genetic variability. Bathyergidae OR genes were classified based on phylogenetic relationships with a range of published OR subgenomes, and compared to orthologous ORs from a number of mammalian species. This data represents the first assessment of OR gene diversity and classification in a subterranean mammal, and is used in subsequent chapters to test hypotheses regarding the evolutionary forces influencing olfactory gene evolution and the role of olfaction in African mole-rats.

University of Cape Town

2.2 Methods

2.2.1 Olfactory Receptor gene isolation and identification

Genomic DNA was extracted from fresh muscle tissue using a standard phenol-chloroform protocol (Sambrook et al. 1989). Species sampled include representative taxa from all currently recognised genera in the *Bathyergidae*: *Bathyergus janetta* (BJ), *Bathyergus suillus* (BS), *Cryptomys hottentotus hottentotus* (CHH), *Cryptomys hottentotus natalensis* (CHN), *Cryptomys hottentotus pretoriae* (CHP), *Fukomys mechowi* (CM), *Fukomys amatus* (CA), *Fukomys anselli* (CAN), *Fukomys bocagei* (CB), *Fukomys damarensis* (CDM), *Fukomys darlingi* (CD), *Georychus capensis* (GC), *Heliophobius argentocinereus* (HA), *Heterocephalus glaber* (HG).

Vertebrate olfactory receptors display a conserved overall structure typical of GPCRs, with variability concentrated across the ligand binding pockets, spanning transmembrane (TM) domains 3-6 (Gaillard et al.2002, Katada et al. 2005). Using the degenerate PCR primers A4/B6 described in Buck and Axel (1991), TM 2-7 from a single *C. damarensis* individual was first amplified to provide a reference sequence for the development of Bathyergid-specific PCR primers; conditions followed those reported in Buck and Axel (1991). PCR products were gel purified using the Wizard® SV Gel and PCR Clean-up System (Promega) and cloned using the pGEM-T-Easy Vector System (Promega). Ligated products were transformed into *Escherichia coli* DH5α $CaCl_2$-competent cells by standard heat-shock treatment (Dagert and Ehrlich, 1979). Insert-containing clones were sequenced using a BigDye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems). Post-sequencing purification was performed using Centrisep Columns (Princeton) and DNA sequences were determined on an ABI 3130 Genetic Analyser using 3130 Genetic Analyser Data Collection software v.5.2.

Because the A4/B6 degenerate primer pair only amplified ORs from a single mole-rat species, bathyergid-specific primers were designed based from the cloned OR sequences from *C. damarensis* (obtained with A4/B6). Bathyergid-specific OR primers target OR TM domains 2-7 (approx. size 645 bps): Bathy-OR1 5'- GCG GAC ATC YGT TTC AC - 3'; Bathy-OR2 5'- GTG ACC ACA GTG TAC ATC – 3'. The

University of Cape Town

Bathy-OR1/Bathy-OR2 primer pair successfully amplified unambiguous PCR products in all 14 mole-rat species, using the following conditions: 95°C for 1 min, 54°C for 3 min, 72°C for 3 min (35 cycles). Each 40 ul reaction contained between 50-100 ng genomic DNA, 2 pmol/ul of each Bathy-OR1 and Bathy-OR2 primers, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1.25 U Super-Therm *Taq* DNA polymerase and 1x corresponding *Taq* reaction buffer. PCR products were gel-purified, cloned and sequenced as described previously. Between 10 and 50 clones were forward- and reverse-sequenced across 1 to 3 individuals for each species, producing a total of 402 OR sequences. Forward and reverse sequences were aligned and checked for ambiguities by eye, then collapsed in Bioedit v7.0.8.0 (Hall 1999), resulting in 201 putative OR sequences.

The 201 putative OR nucleotide sequences were aligned using Clustal W v2.0 (Thompson et al. 1994, Larkin et al. 2007) and translated within Bioedit v7.0.8.0 (Hall 1999); corrections were made by eye. Identical sequences were identified by pairwise comparisons in MEGA v4 (Tamura et al. 2007), resulting in a final data set of 178 unique OR sequences.

University of Cape Town

2.2.2 Recombination test

The role of recombination in generating sequence variation in this dataset, either *in vivo* or *in vitro*, was evaluated by calculating the level of linkage disequilibrium between polymorphic sites as a function of their physical distance, using Rozas et al's (2001) ZZ value in DnaSP4.5. ZZ is calculated as:

$$ZZ=Z_A-Z_{ns}$$

where Z_A is the average linkage disequilibrium (r^2) (Hill & Robertson, 1968) between adjacent polymorphic sites, and Z_{ns} is the average r^2 over all pairwise comparisons (Kelly 1997). The aim of performing this test is to discard the possibility of *in vitro* recombination generating false OR variability, since recombination is established as only a minor source of OR variation *in vivo* (Nei & Rooney 2005).

2.2.3 Assigning sequence identity

A database BLAST search was performed against the nucleotide collection data available on NCBI (www.ncbi.nlm.nih.gov) to assign identity to both nucleotide and amino acid sequences, using the 'highly similar' search option. In the majority of searches, high sequence homology was found between the Bathyergidae and other mammalian OR genes. Where high similarity was not evident, the 'somewhat similar' option was used and the resulting best scores were always attributable to OR genes. Secondly, the NCBI tool designed to identify 'conserved motifs' in query protein sequences was used, and led to the detection of known GPCR domains among all protein sequences of our dataset. An attempt to annotate nucleotide sequences using Blast2GO was also done (Conesa et al. 2005, Götz et al 2008). Annotation, however, was inconsistent due to the discrepancy in OR nomenclature across published sequences. For example, mammalian OR genes are traditionally classified into Class I and Class II (Glusman et al. 2000a, Zhang & Firestein 2002); however, Class I corresponds to groups α and β, and Class II to group γ, in another widely accepted nomenclature which was designed for vertebrates in general (Niimura and Nei 2005, 2006, Niimura 2009). An explanatory nomenclature, similar to that recently used in Hayden et al. (2010), is therefore used to describe OR sequences in this dataset (Appendix I.2).

2.2.4 Identification of pseudogenes

Following Steiger et al. (2009) OR sequences were classified as pseudogenes if they contained stop codons or frame-shift mutations that disrupted the overall receptor structure. Sequences that translated into putatively functional OR genes, but that differed in length, were considered to be functional only if they maintained the known features of ORs (e.g. the MAYDRFVAIC and KAFSTCASH motifs in TM domains 3 and 6, respectively – see Appendix II.1 for a protein alignment of the putatively functional OR sequences), and if the variability mapped to the ligand-binding pockets of ORs (Man et al. 2004, Katada et al. 2005).

University of Cape Town

2.2.5 Olfactory Receptor molecular structure

A cartoon of a typical Bathyergidae OR was constructed with reference to Katada et al.'s (2005) molecular model of the mouse mOR-EG receptor (Kajiya et al. 2001).

In order to facilitate visual sequence comparisons between Bathyergid ORs and the mouse mOR-EG receptor (Kajiya et al. 2001), a protein sequence 'logo' was created in LogoBar – 0.9.12 (Pérez-Bercoff et al. 2006) using the amino acid alignment of putatively functional Bathyergidae ORs. The logo displays a consensus sequence for the amino acid alignment, and a visual representation of sequence diversity based on amino-acid frequencies per site (Figure 2.3). Highly conserved residues, common to both the Bathyergidae receptors and mOR-EG, were identified and used to orientate the molecule and generate a cartoon model for the Bathyergid OR.

University of Cape Town

2.2.6 Identification of alleles

In order to identify allelic variants of OR genes, pairwise comparisons were performed across all unique sequences in our dataset using MEGA v4 (Tamura et al. 2007). Allelic pairs based on the pairwise comparison matrix generated in MEGA were then identified using *alleles.R* (R. Gaujoux, unpublished) developed in R (R Development Core Team, 2008, http://www.R-project.org). The criteria for allele identification described by Kishida (2008) were adapted to bathyergids using the following cut-off limits: within a species, sequences that shared 99% sequence similarity were considered to be alleles of the same gene; across species, the cut-offs were 98% within the same genus and 96% across bathyergid genera. Single base-pair differences as well as two base-pair differences were never considered allelic variants, but were assumed to represent identical sequences due to PCR or sequencing errors. Similarly, when two allelic variants shared more than 99% sequence similarity across species (i.e. between 3-5 base pair differences), they were considered to represent identical alleles (i.e. trans-species polymorphisms, see discussion 2.4.5 below). When more than two putative alleles of the same OR gene were found in an individual given the defined cut-offs, two copies of that particular gene were assumed to be present**.** Similarly, when two presumed alleles were of different functional status i.e. one putatively functional and one pseudogene, they were considered to belong to two different OR genes, the result of a duplication event followed by pseudogenisation.

Whenever the % sequence similarity led to ambiguous results e.g. when transitivity was not applicable (A=B, B=C but A≠C, with '=' meaning 'alleles' based on sequence similarity), phylogenetic relationships (see description below) were used to allocate alleles to different OR genes. Once identified, alleles of the same OR gene were collapsed down to a single representative sequence for each putative gene, that was used in subsequent analyses (Appendix I.3). In other words, only one representative allelic sequence was arbitrarily chosen to represent a particular OR gene.

2.2.7 OR evolutionary relationships within Bathyergidae

Evolutionary relationships among Bathyergidae OR genes were explored using a maximum likelihood (ML) tree (Felsenstein 1981) based on the general time-reversible model (GTR, Tavaré 1986) constructed in MEGA v5 (Tamura et al. 2011). To find the most appropriate model for maximum likelihood (in this case GTR), a jModeltest analysis was performed on the dataset (Posada 2008). Tree topology was inferred using all unique African mole-rat OR sequences identified, together with three non-OR GPCR genes to root the tree (Appendix III.1); robustness of the tree topology was tested using 1000 bootstrap replicates (Felsenstein 1985). The resulting tree was used in combination with the pairwise comparison matrix to determine allelic relationships amongst sequences. If sequence similarity led to uncertain allelic allocation, alleles were considered to be sister taxa in the phylogenetic tree (Appendix III.1). A further ML tree (Felsenstein 1981) was then constructed (GTR, 1000 bootstrap) (Tavaré 1986, Felsenstein 1985), using only a single representative sequence for each putative OR gene (Figure 2.5 and Appendix III.2).

2.2.8 Phylogenetic relationships between Bathyergid and Mammalian OR genes

A recent study by Hayden et al. (2010) used a combination of sequence similarity and phylogenetic criteria for OR gene classification in the most comprehensive survey on mammalian ORs. Hayden et al.'s dataset included the entire OR subgenomes of 50 mammalian species, consisting of 50,000 OR sequences circa. Of these, ~2,000 OR genes were characterised using a PCR-based method across 18 mammalian species,

University of Cape Town

and ~48,000 ORs were recovered from *in silico* data mining searches. All the traditional OR families (Glusman et al. 2000a) were recovered in Hayden et al.'s study (2010), and the majority of them were found to be monophyletic. Only the following three OR families appeared polyphyletic: families 2 and 13, families 1, 3 and 7, and families 5, 8 and 9.

Of the 50 000 OR sequences in Hayden et al.'s dataset (2010), only the 2000 that were characterised via a PCR-based approach were publicly available, and have been used in this section. The available OR sequences from Hayden et al.'s study (2010), representing entire OR repertoires of 18 mammalian species, were downloaded from the NCBI data bank (www.ncbi.nlm.nih.gov).

One or two representative sequences per species for each of the 17 OR families present in their dataset were aligned together with the Bathyergid OR dataset, using the online Clustal W alignment tool from the European Bioinformatics Institute (available at www.ebi.ac.uk). Aligned sequences were then imported into Bioedit v7.0.8.0 (Hall 1999) and corrections to the alignment were made by eye. A maximum likelihood (ML) tree obtained under the Tamura-Nei substitution model (Felsenstein 1981, Tamura and Nei 1993) was constructed in MEGA v5 (Tamura et al. 2011) after 1000 runs of bootstrapping (Felsenstein 1985). The most appropriate model for maximum likelihood (Tamura-Nei, Tamura and Nei 1993) was chosen using jModeltest (Posada 2008). Phylogenetic analysis based on all nucleotide sites included 312 representative sequences from all OR gene families across 18 different mammalian species, as well as the 119 Bathyergidae ORs. All Bathyergidae sequences clustered together with known Family 7 ORs (Hayden et al. 2010).

A database BLAST search was performed with the original Rat OR sequences described by Buck and Axel in their seminal paper (1991), against the nucleotide collection data available on NCBI (www.ncbi.nlm.nih.gov). Among the 18 sequences characterised by Buck and Axel (1991), three OR families were represented, namely families 1, 6 and 7. Therefore, Buck and Axel's (1991) A4/B6 degenerate primers may have been biased to only amplify genes from those three OR families. Alternatively, Buck and Axel's sequencing effort (i.e. number of clones sequenced) was insufficient to detect OR genes from other families, which may have been

University of Cape Town

amplified by A4/B6. The bias in Bathyergid-specific primers towards amplifying family 7 genes exclusively is consistent with both these explanations, and emerges as a feature of the Bathy-OR1/Bathy-OR2 primer pair designed in this study.

A new database BLAST search was performed with all African mole-rat OR genes against the nucleotide collection data available on NCBI (www.ncbi.nlm.nih.gov), in order to confirm OR family identity by sequence similarity criteria. Family information was included in Bathyergidae sequence nomenclature (Appendix I.2).

2.2.9 Analysis of orthologous OR genes across species

All OR sequences belonging to Family 7 from Hayden et al's (2010) dataset, representing the entire Family 7 OR subgenome of 18 different mammalian species, were aligned with the 119 mole-rat OR genes using the online Clustal W tool like previously described (www.ebi.ac.uk). Aligned sequences were corrected by eye in Bioedit v7.0.8.0 (Hall 1999) and a maximum-likelihood tree (Tamura-Nei) was constructed in MEGA v5 (Tamura et al. 2011) and tested using 1000 bootstraps samples (Felsenstein 1985). Again, the Tamura-Nei (Tamura and Nei 1993) substitution model was chosen based on the results of a jModeltset analysis (Posada 2008). Positions containing alignment gaps were eliminated from the pairwise sequence comparisons (pairwise deletion option), resulting in 805 nucleotide positions in the final dataset.

University of Cape Town

2.3 Results

2.3.1 Isolation of olfactory receptor loci in African mole-rats

The Bathy-OR1/Bathy-OR2 primer pair designed in this study resulted in unambiguous amplification of OR loci in all the 14 African mole-rat species analysed. On average, the loci targeted were ~640 bps long, spanning TM domains 2 to 7 of OR genes (the full nucleotide alignment is reported in Figure 1 – Supplementary Material

on CD). Between 10 and 50 clones were successfully sequenced across 1-3 individuals for each species, resulting in 201 sequences that corresponded to 178 unique OR nucleotide sequences.

As with other species studied to date recombination is not a significant mechanism for the generation of sequence variability across mole-rat OR loci (Nei and Rooney 2005). The significance of pairwise associations between polymorphic sites (linkage disequilibrium) was assessed using a chi-square test and Rozas et al's (2001) ZZ statistic was calculated as a measure of overall linkage disequilibrium between polymorphic sites. The ZZ statistic is predicted to have large positive values with increasing recombination and was not significant across the dataset (ZZ=0.006). This result suggests that *in vitro* recombination, resulting from PCR recombination artefacts (Meyerhans et al. 1990), has not occurred. Furthermore, this result is consistent with the widely accepted idea that variability across OR genes is predominantly the result of gene duplication events and nucleotide substitution driven by positive selection, rather than recombination (Nei and Rooney 2005).

A BLAST search using the NCBI GenBank database (www.ncbi.nlm.nih.gov) confirmed the sequence identity as ORs for all sequences in the dataset. A 'conserved domains' search revealed the presence of typical GPCRs features in all sequences (Terakita 2005), whilst known OR motifs were confirmed by eye from the amino acid alignment (Zhang & Firesetin 2002, Godfrey et al. 2004).

University of Cape Town

2.3.2 Pseudogenes

Consistent with published studies, mole-rat OR sequences were considered to be pseudogenes if they contained disruptions to the 7TM receptor structure; these disruptions included stop codons and frameshift mutations (Steiger et al. 2009). Using these criteria, 97 of the 178 Bathyergid OR sequences were classified as pseudogenes. This may be an underestimation of the extent of non-functional OR genes, because mutations mapping outside the partial coding region amplified (TM 2-7), or in

promoter regions would not be detected (Gilad et al. 2004, Rouquier et al. 2000). The extent of pseudogenisation in the mole-rat OR repertoire characterised is noteworthy, and is the subject of further discussion and analysis in the next two chapters.

2.3.3 Global analysis of Bathyergid OR sequences

To gain a comprehensive view of the amino acid and nucleotide diversity across the data set, protein sequence 'logos' were generated using the full Bathyergid OR amino acid alignment (Figure 2.3). Although the large number of genes isolated in the data set tends to reduce the degree of conservation among amino-acid residues (Zhang and Firestein 2002), the characteristic features of OR sequences can still be observed (e.g. MAYDRFVAICH in TM3, or KAFSTCGSH in TM6, Zhang & Firesetin 2002, Katada et al. 2005). Predicted locations for transmembrane domains (TM2-TM7) are shown in Figure 2.3.

University of Cape Town

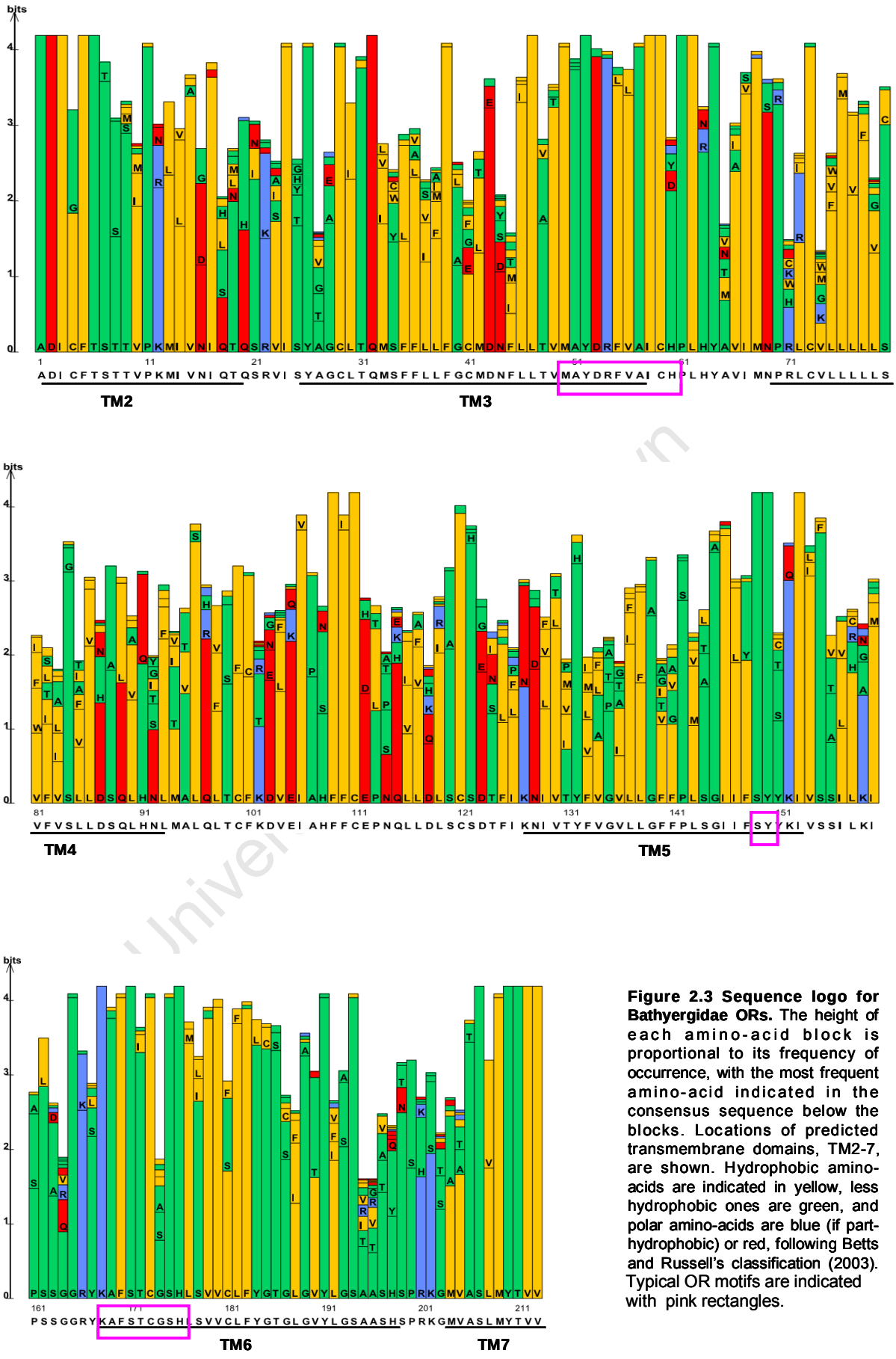


Figure 2.3 Sequence logo for Bathyergidae ORs. The height of each amino-acid block is proportional to its frequency of occurrence, with the most frequent amino-acid indicated in the consensus sequence below the blocks. Locations of predicted transmembrane domains, TM2-7, are shown. Hydrophobic amino-acids are indicated in yellow, less hydrophobic ones are green, and polar amino-acids are blue (if part-hydrophobic) or red, following Betts and Russell's classification (2003). Typical OR motifs are indicated with pink rectangles.

A cartoon showing the distribution of amino acid diversity across the Bathyergidae ORs isolated in this study was generated based on Katada et al.'s (2005) molecular model of the mouse mOR-EG receptor (Figure 2.4). The topological distribution of conserved and variable sites (Figure 2.4.a) in the mole-rat receptor is analogous to that of mOR-EG (Katada et al. 2005), with 73 % of highly conserved residues shared and 88 % of variable residues occupying the same locations.

In an attempt to identify the amino-acid sites involved in odorant-binding, Man et al. (2004) hypothesised that such sites would be highly conserved among orthologous ORs, but polymorphic across paralogous ORs. A comparison of sets of OR orthologs and paralogs in human and mouse allowed Man et al. (2004) to identify 22 predicted odorant-binding amino-acid sites, spanning TM 2-7 (but concentrated in TM 3-6). Katada et al. (2005), on the other hand, investigated the ligand-binding domain of a mouse OR via a series of site-directed mutants and ligand docking simulations and found nine candidate amino-acid positions involved in odour-binding, all in TM 3-6, four of which correspond to the ones described by Man et al. (2004). Together, these two studies identify 27 amino-acid sites as good candidates for the odorant-binding domain of ORs (Katada et al. 2005, Man et al. 2004), 26 of which fall in the portion of Bathyergidae ORs sequenced in this study (Figure 2.4.c).

High levels of both nucleotide and amino acid sequence polymorphism were detected in mole-rat OR sequences, and variability appears to be concentrated in the region between TM3 and TM6 (Figures 2.3 and 2.4), which corresponds to the predicted core of the ligand binding pocket of ORs (Katada et al. 2005, Man et al. 2004). In particular, 19 of the 26 amino acid residues predicted to be involved in ligand-binding, according to the combined results of Man et al. (2004) and Katada et al. (2005), are variable in Bathyergids (Figure 2.4.c). If residues in TM domains 2 and 7 are excluded, 83% of the alleged odorant-binding sites in mole-rat ORs are polymorphic, consistent with a role in odorant recognition (Man et al. 2004).

University of Cape Town

Figure 2.4

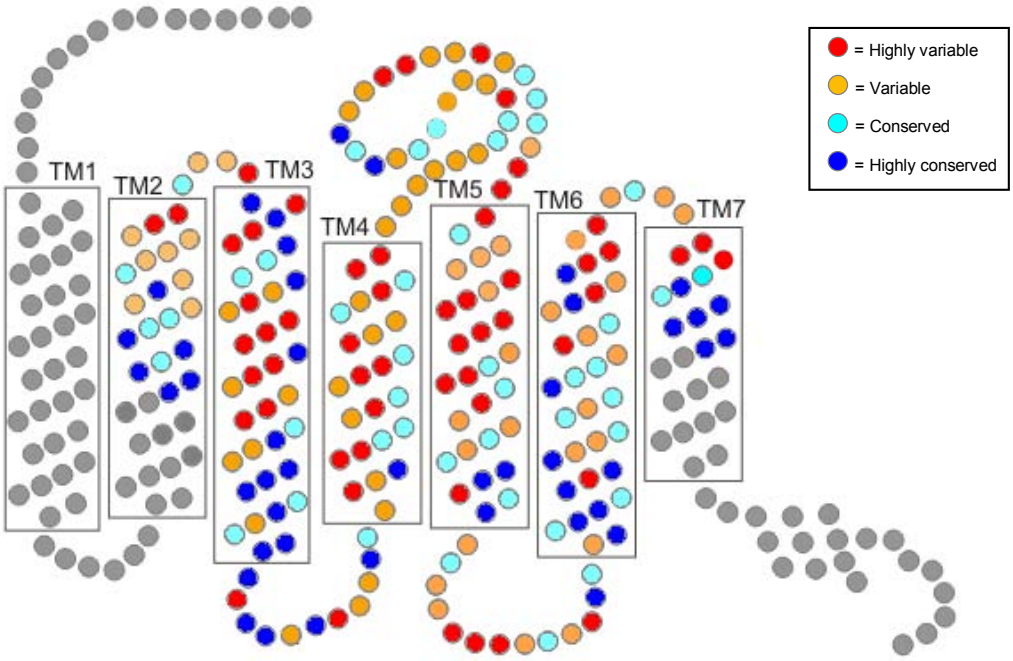


Figure 2.4.a Variability across functional ORs. The number of different amino acids per site determines the level of variability: ≥ 5 = highly variable, 3-4 = variable, 2 = conserved,1= highly conserved. Grey dots correspond to non amplified amino-acid positions

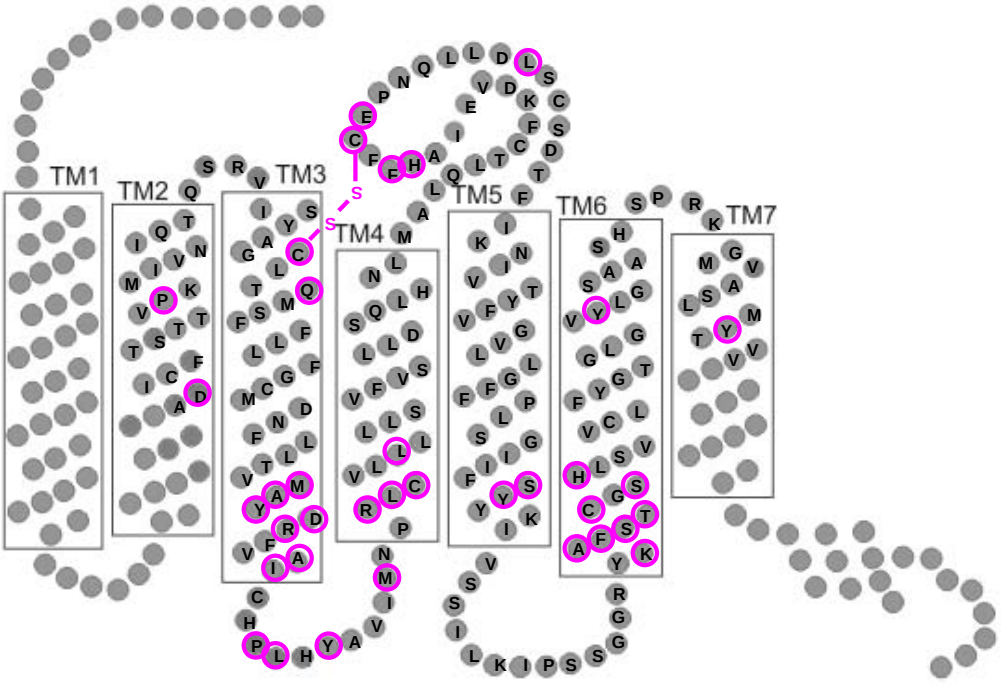


Figure 2.4.b Comparison of sequence conservation between mOR-EG (mouse, Kajiya et al. 2001) and Bathyergidae OR genes. The sequence displayed is the consensus sequence generated from the protein alignment of putatively functional OR genes. Amino-acid positions that are highly conserved in both mole-rat ORs and mOR-EG are circled in pink and were used to orientate mole-rat receptors according to the mouse model (Katada et al. 2005). Grey dots correspond to non amplified amino-acid positions.

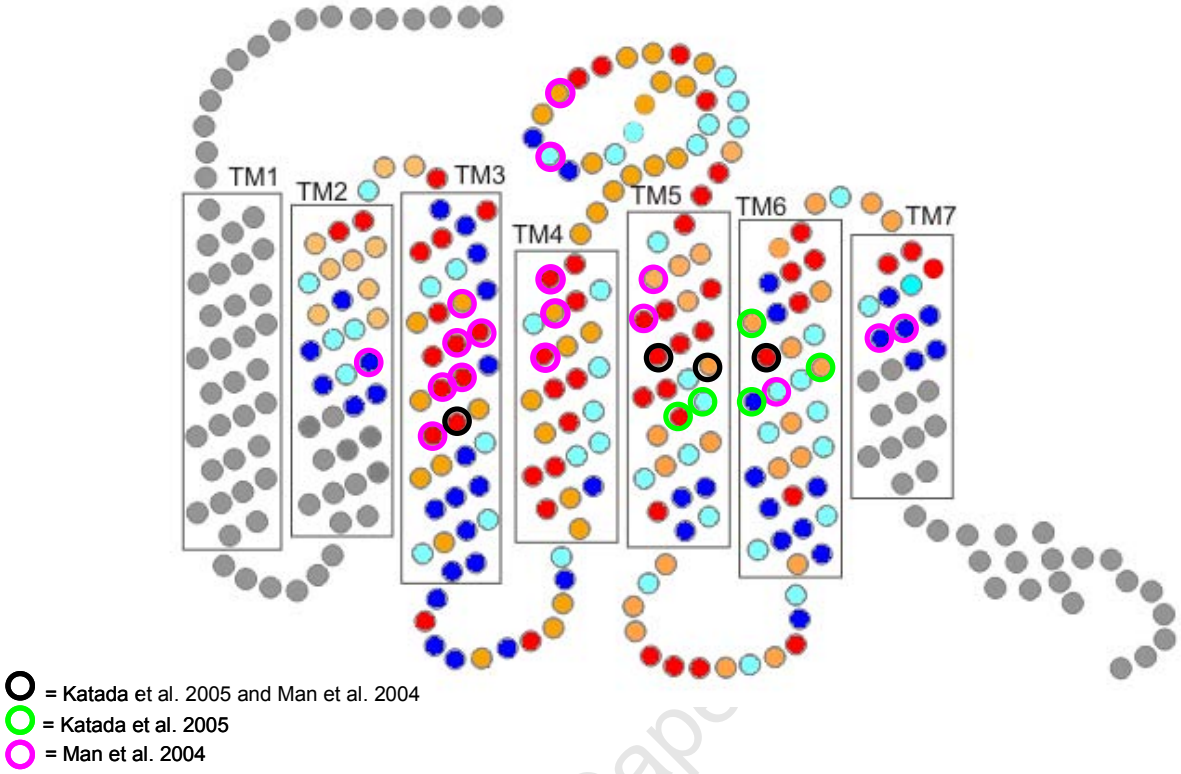


Figure 2.4.c Amino-acids involved in odorant-binding. Circled amino-acid positions are predicted to be involved in the odorant-binding site of ORs according to Man et al. 2004 and/or Katada et al. 2005. Amino-acids are colour-coded as per Figure 2.3.a based on sequence variability.

University of Cape

2.3.4 Estimating the number of unique mole-rat OR genes

Two possible sources of intra-locus variation need to be taken into consideration when estimating the number of OR genes: true allelic variability and artefacts resulting from *Taq* polymerase errors (Whinnet & Mundy, 2003). The choice of standard *Taq* as opposed to a specific proof-reading *Taq* was informed by the results of previous studies conducted in our laboratory, which revealed high fidelity of the regular enzyme under analogous circumstances (unpublished data). In this context, it is important to note that the unlikely occurrence of PCR biases would be comparable across the data sampled. To further minimise sequencing artefacts, each OR clone was forward- and reverse-sequenced and the two sequences were combined to form contigs only if no sequencing ambiguities were found. Furthermore, a very conservative method of allele identification, following Kishida (2008), was applied to the dataset (see Methods), reducing the chance of falsely identifying signatures of genetic variability among the OR genes characterised in this study.

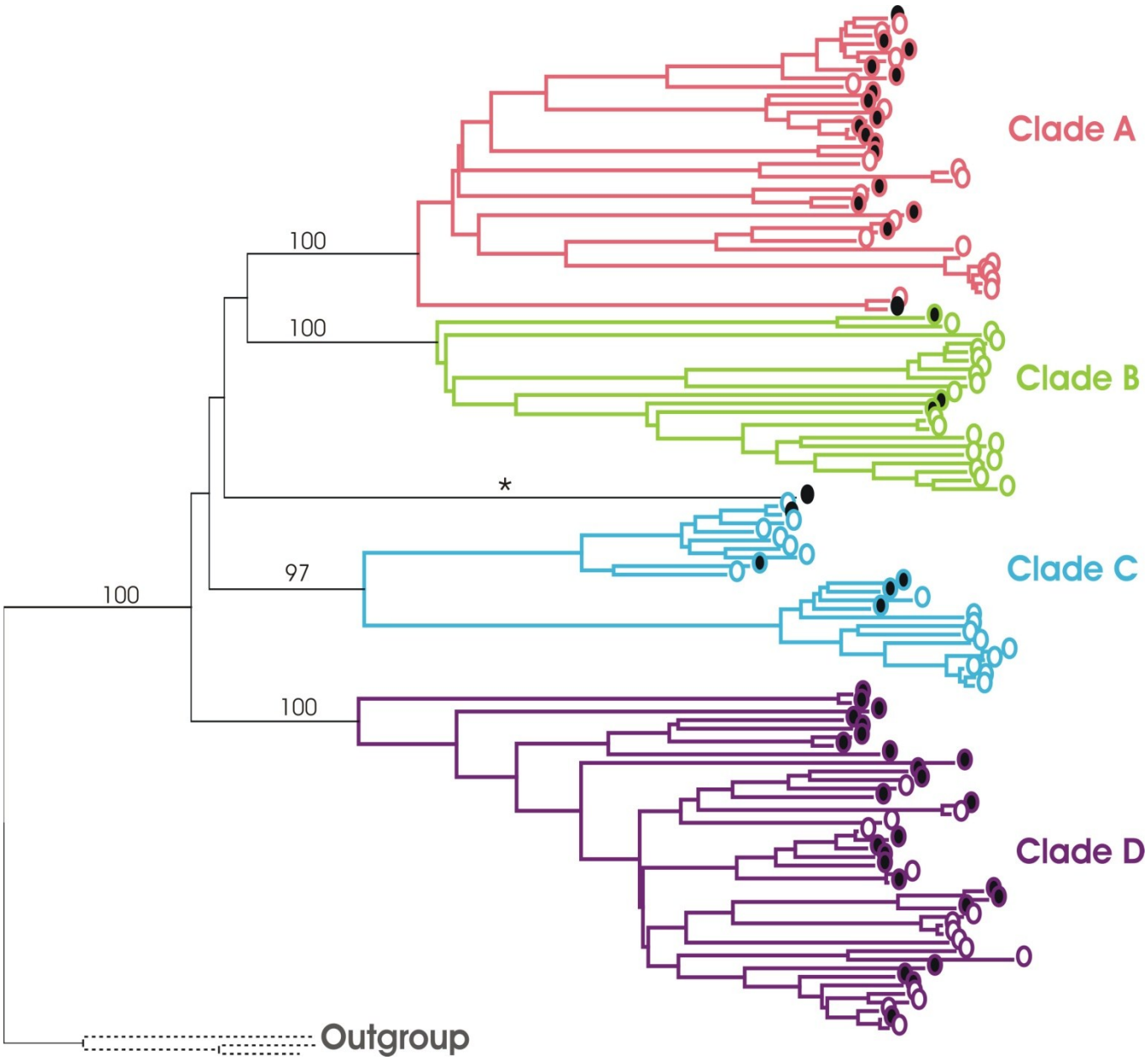
After allelic variants were merged, a total of 119 unique OR genes were identified from the original pool of 178 OR gene candidates. These unique genes include 51 putatively functional ORs and 68 OR pseudogenes. Interestingly, alleles of the same OR gene (as well as identical alleles) were detected on several occasions across mole-rat species (Appendix 1.4). Importantly, this result suggests that some degree of OR diversification may have preceded speciation in Bathyergidae, consistent - in the case of identical alleles being present across species - with a scenario of 'trans-species evolution' which will be discussed in the next section of this chapter (Klein et al. 1998, Takahata and Nei 1990).

2.3.5 Phylogenetic relationships among mole-rat OR sequences

The neighbour-joining analysis (Kimura-two-parameter) using the 178 African mole-rat OR sequences characterised in this study is represented in Appendix III.1. The tree reveals four well supported clades of closely-related OR genes (bootstrap support ≥97), indicated as clades A-D, as well as an isolated gene, namely BJ4_A12, which constitutes a sister lineage to clades A and B. The same tree topology is recovered in Figure 2.5, where only a single representative sequence appears for each putative OR gene (the full tree comprising gene names can be found in Appendix III.2).

University of Cape Town

Figure 2.5 Bathyergidae OR gene tree Simplified schematic view of the maximum likelihood tree (GTR, 1000 bootstrap) constructed using a single representative sequence for each putative Bathyergid OR gene; three rhodopsin-like GPCRs were used to root the tree (accession numbers NP_001287.2, NP_005292.2, NP_037014.2). The four main OR clades - A-D - are indicated in different colours, and bootstrap support values are reported for each clade; only one isolated gene falls out of these clades and is indicated with an asterisk. Black filled dots at branch tips represent the putatively functional OR genes; empty dots represent OR pseudogenes. An alternative view of this tree, inclusive of gene names, can be found in Appendix III.2.



Across clades A-D OR genes do not appear to cluster in a species-specific way, but rather they cluster based on sequence similarity according to their respective OR genetic lineages. In fact, sequences in each clade share functional motifs across the ligand-binding site of ORs (full alignment of the putative odorant-binding sites across Clades A-D can be found in Appendix II.2). The consensus sequences for the odorant-binding sites across TM domains 3-6 of Clades A-D are reported in Figure 2.6.

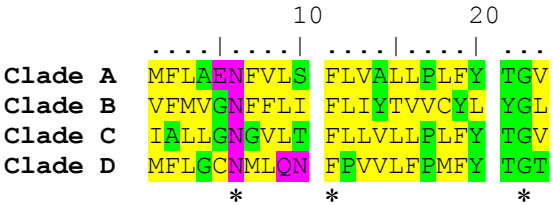


Figure 2.6 Amino-acid sites involved in odorant-binding across TM 3-6 in clades A-D. Conserved sites are marked with an asterisk. Hydrophobic amino-acids are highlighted in yellow (very hydrophobic) and green (less hydrophobic), while hydrophilic amino-acids are highlighted in fuchsia following Betts and Russell (2003).

Of 23 amino-acid positions involved in odorant-binding across TM3-6 (Man et al. 2004, Katada et al. 2005), only three are conserved across clades A-D, whilst the remaining 20 odorant-binding sites display clade-specific motifs, suggesting that ORs in each clade have different binding properties. Furthermore, it is interesting to note the striking prevalence of hydrophobic amino-acids (yellow and green in Figure 2.6) across the putative ligand-binding domain of ORs in all clades. Importantly, this finding is consistent with Katada et al.'s (2005) hypothesis that the interaction between ORs and odorant ligands occurs primarily via hydrophobic and van der Waals interactions (Parsegian 2006).

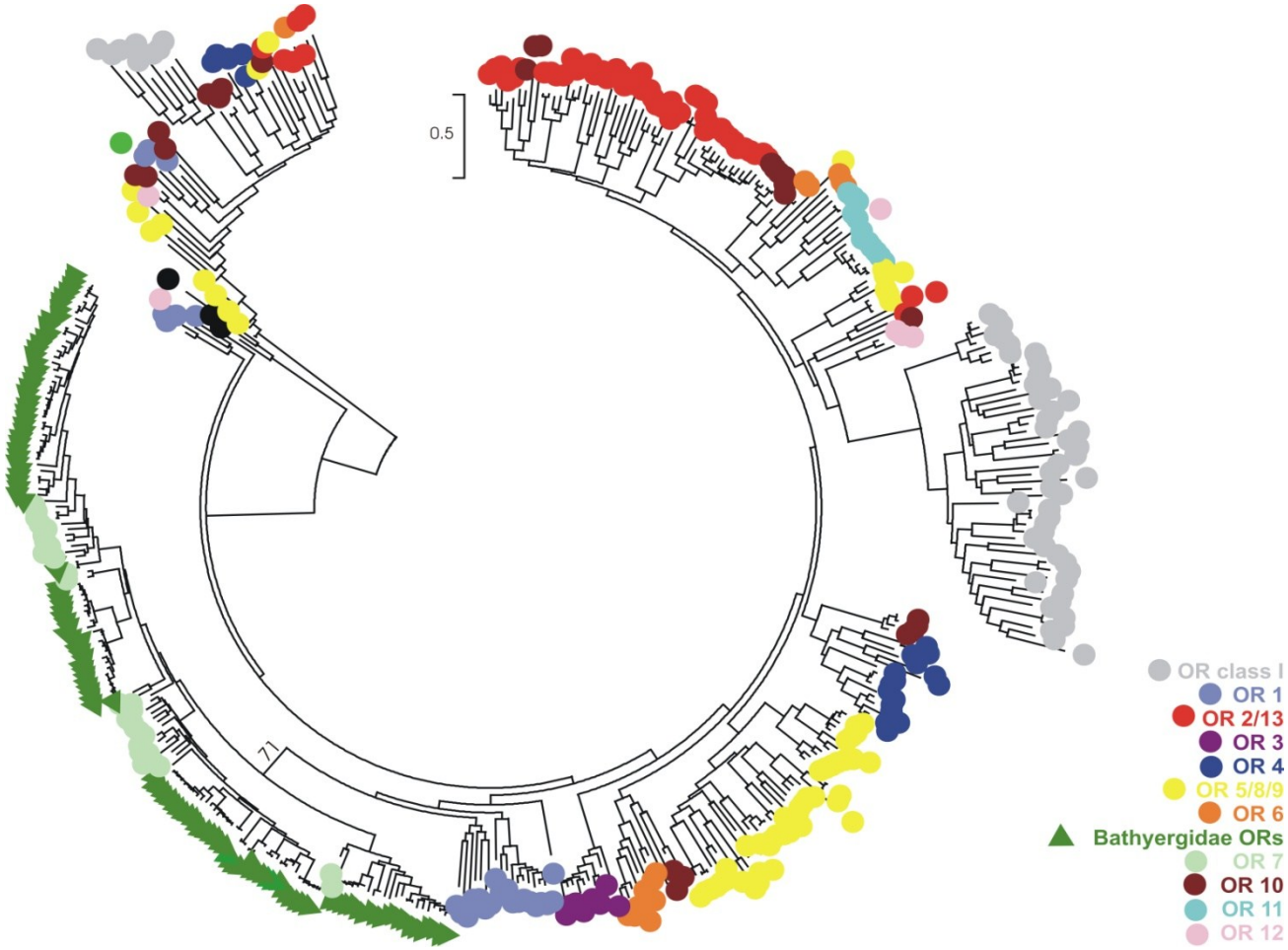
2.3.6 Classification of Bathyergid OR genes

In order to identify which portion of the OR subgenome was amplified by the Bathy-1/Bathy-2 primer pair, phylogenetic relationships between mole-rat OR genes and representative sequences from entire OR repertoires of 18 different mammalian species were inferred (Hayden et al. 2010) using the maximum likelihood method (Tamura-Nei substitution model) (Felsenstein 1981, Tamura and Nei 1993) with 1000 bootstrap replicates (Felsenstein 1985). The resulting tree reveals strong support for

University of Cape Town

Bathyergidae ORs clustering together with Family 7 OR genes from a number of mammalian species (Figure 2.7).

Figure 2.7 Mammalian OR Family structure Maximum likelihood tree obtained with Tamura-Nei substitution model (1000 bootstrap) using representative sequences of all OR Families from the available Mammalian database (Hayden et al. 2010), together with the Bathyergidae OR genes characterised in this study. OR families are colour-coded as reported on the right. All Bathyergidae ORs appear to cluster together with mammalian Family 7 OR genes (indicated in green, together with the bootstrap support value for that branch).



The direct BLAST search conducted against the nucleotide collection data available on NCBI (www.ncbi.nlm.nih.gov) further supports the above result, validating OR family identity using sequence similarity. For every Bathyergidae OR gene isolated in this study, the best BLAST hits, based on Hayden et al.'s (2010) OR family nomenclature, were listed as family 7 ORs.

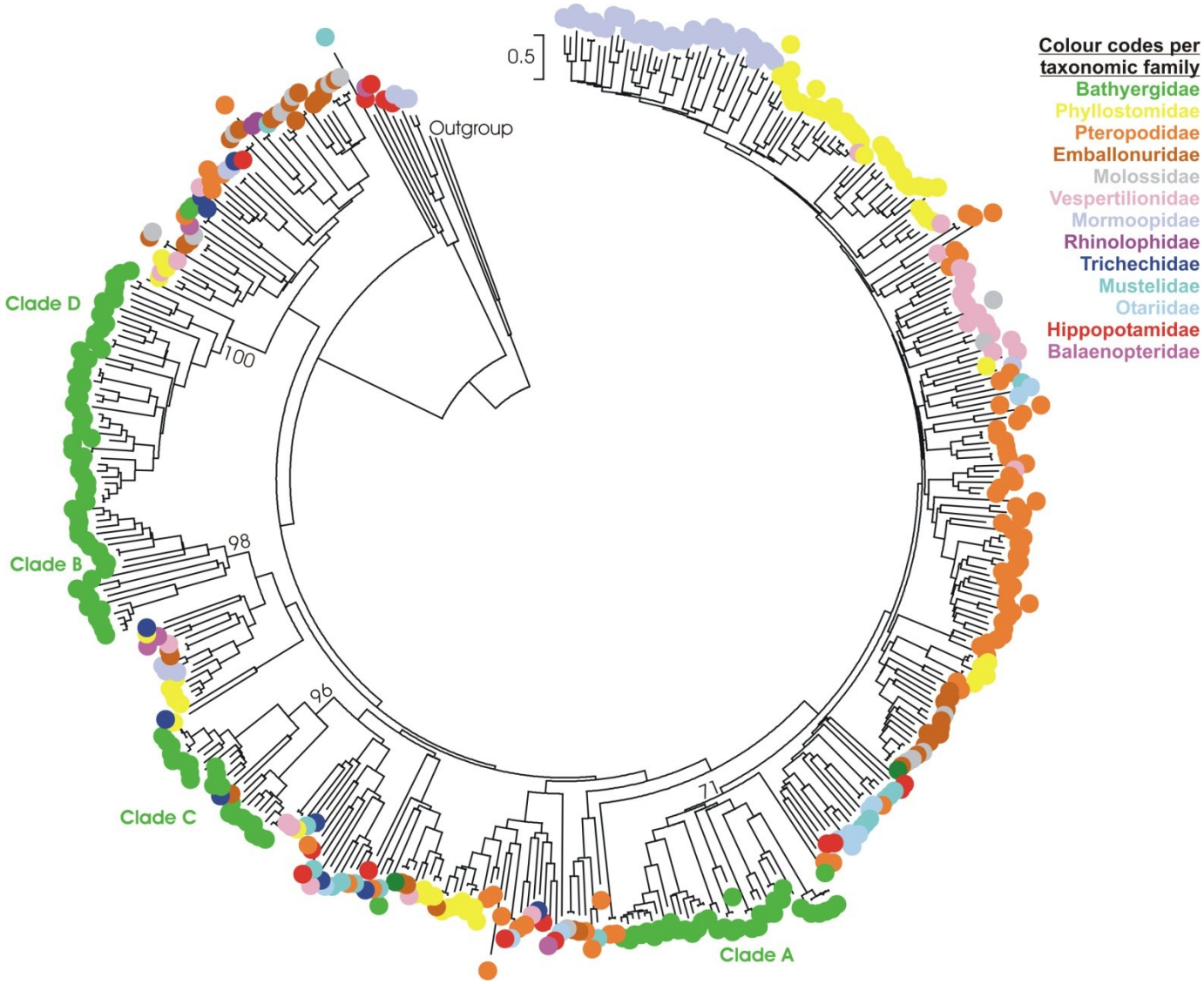
Family 7 OR genes represent a polyphyletic family of Class II OR genes in mammals, and are classified as part of the larger grouping of families 1/3/7 (Hayden et al. 2010). OR genes from families 1 and 3, however, appear to group independently from family 7 in strongly supported clades in our tree (Figure 2.7 and Figure 2 in Supplementary Material on CD), and appear to be more distantly related to the mole-rat OR subgenome captured in this study.

2.3.7 The evolution of Family 7 OR genes in African mole-rats

The evolution of OR7 Bathyergidae genes was inferred by phylogenetic analyses of all the available mammalian OR7 sequences from Hayden et al.'s (2010) dataset (Figure 2.8, and Figure 3 - Supplementary Material on CD - for the same tree inclusive of gene and species information).

University of Cape Town

Figure 2.8 Mammalian OR7 gene tree. Maximum likelihood tree (Tamura-Nei, 1000 boostrap) constructed with all known mammalian OR7 genes (Hayden et al. 2010). Each dot corresponds to an OR gene belonging to family 7; ORs from different taxonomic families are colour-coded as indicated on the right of the figure. Rhodopsin-like non-OR GPCRs are used as an outgroup (accession numbers NP_001287.2, NP_005292.2, NP_037014.2). Bathyergidae ORs from clades A-D are indicated in green; bootstrap values are reported for the main bathyergid clades.



Again, African mole-rat OR genes cluster into four strongly supported clades (Figure 2.8), which correspond to clades A-D in the Bathyergidae phylogenetic tree (with the exception of two genes, CA3_B4 and CAN3_A5, which appear outside Clade D) (see Appendix III.2 and Figure 3 in Supplementary Material on CD for comparison). Interestingly, clades A, B and D appear to be Bathyergidae-specific clades, whilst clade C displays OR7 genes from other mammalian species, but splits in two strongly supported lineage-specific clades of Bathyergidae ORs. Other family-specific clades are highlighted in Figure 2.8 by a colour-coded classification of mammalian OR7 genes, according to the taxonomic lineage of provenance.

University of Cape Town

2.4 Discussion

This chapter presents the first evaluation of a sub-set of OR genes in a family of subterranean rodents, the Bathyergidae. A PCR-based approach was used to isolate and characterise 178 unique OR sequences across 14 species of African mole-rats, and corresponded to the identification of 119 novel genes. The robustness of the experimental procedure was ensured by a series of forethoughts regarding the type of *Taq* polymerase used, the possibility of sequencing errors, and that of *in vitro* recombination in generating false genetic variability. Additionally, in order to avoid overestimation of OR diversity, highly conservative criteria were applied in the process of allocating alleles to Bathyergid OR genes. In order to classify the Bathyergidae OR gene repertoire identified in this study, phylogenetic inference between mole-rat ORs and a range of mammalian OR subgenomes was used. A single family of Bathyergid OR genes appears to have been preferentially captured by the mole-rat specific primer pair designed in the study, namely OR family 7, OR7 (Glusman et al. 2000a, Hayden et al. 2010). Interesting features of Bathyergid OR7 genes include the presence of functional polymorphisms presumably translating into diverse binding properties within the gene family, as well as the presence of trans-species polymorphisms suggesting an ancient origin for some aspects of OR diversification in the Bathyergidae (Klein et al. 1998).

University of Cape Town

2.4.1 Characterisation of Bathyergidae ORs: notes on methodology

a) Sampling effort

The limiting factor in a PCR-based approach for the characterisation of a multi-gene family lies in the sampling effort, which in this case corresponds to the number of Bathyergid OR clones that were sequenced. If, for instance, out of 100 clones sequenced, only 15 unique sequences recur multiple times, one could confidently say that all the genes that could be amplified by the primers used are likely to have been detected. On the other hand, if out of 100 sequences, 90 are unique, it is probable that the primers used have not fully explored the potential diversity in the data set (Hayden et al. 2010). In this study, out of 201 Bathyergid OR clones sequenced, 178 OR sequences were unique, suggesting that the primers used could have amplified more

OR genes than have been recovered. Nevertheless, the fact that the Bathyergid OR genes characterised here belong to a single OR family, namely OR7, provides a highly informative dataset with which to further explore the evolution and function of OR genes in African mole-rats (see discussion below).

b) The extent of OR pseudogenes and functional genes

In comparing OR subgenomes across species, the proportions of pseudogenes found are of great importance because they are thought to i) reflect the degree of olfactory aptitude of a species, and ii) reflect aspects of the evolutionary history of ORs (Gilad et al. 2004, Kishida et al. 2007). Among the 119 OR genes in this dataset, 51 appear to be putatively functional, translating into proteins that display all the structural motifs of both GPCRs and ORs, whilst 68 have been classified as pseudogenes. Although the proportion of pseudogenes might be underestimated, since mutations mapping outside the area characterised here would not be revealed (Gilad et al. 2004, Rouquier et al. 2000), it is possible that the conservative methods used to identify alleles may have reduced the estimated number of putatively functional genes. The ratio of functional OR genes:pseudogenes is further explored in Chapter 3 in the context of the evolutionary forces shaping OR diversity and functionality across the Bathyergidae.

c) OR phylogenetic inferences

The dataset used for the phylogenetic comparisons performed in this chapter comprise a variety of non-model mammalian OR repertoires, ranging from bats to aquatic mammals. The analysis was based on the dataset generated by Hayden et al. (2010), which took into consideration all the publicly available sequence information on OR genes, including rodents, for an overall OR family classification across Mammals. As mentioned in this chapter's introduction, the same 17 OR gene families are found to occur across all the mammalian species examined (Hayden et al. 2010). Of the 50 000 OR sequences in Hayden et al.'s dataset (2010), only the 2000 that were characterised via a PCR-based approach were made available. Unfortunately, information on the rodent OR genes used in Hayden et al.'s (2010) study was not publicly available. Although it could be argued that OR information from rodent species would be more

University of Cape Town

suited to establish sequence orthology with Bathyergidae ORs, I am confident that the results presented here are accurate since the main purpose was to categorise Bathyergid ORs into known gene families. A more integrated dataset, including OR information from a range of terrestrial mammals, is used in Chapter 4.

2.4.2 Bathyergidae OR classification

OR family 7 genes, classified as Class II ORs (group γ), represent one of the largest OR families in mammals (Glusman et al. 2000a, Niimura and Nei 2005b, Hayden et al. 2010), and are thought to have diverged after the Placental-Marsupial split of early Mammals (~147 MYA according to a recent molecular dating based on 66 genes and over 2000 mammalian species - Bininda-Emonds et al. 2007, Kishida 2008). In humans, high levels of gene duplication have resulted in OR7 being the largest family of the OR subgenome, occurring as OR7-specific clusters scattered across a number of genomic locations (Glusman et al. 2001). Although the function of OR7 is poorly understood, some genes belonging to OR7 are crucial in recent mammalian evolution, e.g. OR7D4 in humans which detects androstenone and androstadienone (Keller et al 2007). Interestingly, these two compounds were classified as human 'pheromone candidates' after they were found to influence both brain function and, more recently, endocrine balance in humans (Jacob et al. 2001a, Wyart et al. 2007). Androstadienone, in particular, is a steroid found in male sweat, saliva and semen and is known to alter mood, physiological arousal and brain activity in both a sex-specific and sexual orientation-specific manner (Grosser et al. 2000, Jacob et al. 2001b, Savic et al. 2001, Bensafi et al. 2004a, Lundstrom and Olsson, 2005). Wyart et al. (2007) recently discovered that androstadienone fulfils an additional key requirement for being considered a pheromone: it is able to influence hormonal balances in conspecifics by inducing high levels of salivary cortisol in women that are exposed to its smell. Cortisol is commonly known as 'the stress hormone' and is found to influence both mood and sexual arousal in humans (Brown and Heninger 1975, Hamilton et al. 2008). Wyart et al.'s finding (2007) is particularly interesting in that, not only does it support the existence of human pheromones but it also suggests that in a species like *Homo sapiens* where the VNO is considered to be a 'nonchemosensory vestige' (Bhatnagar and Smith 2010), pheromonal communication may still occur, possibly mediated by ORs in the MOE. The existence of human

University of Cape Town

pheromones is a highly contentious issue and a matter of strong debate (see Wysocki and Preti 2004), however the above examples of OR7D4 are the only published research suggesting a putative role for the OR7 family.

A recent study by Smith et al. (2007) revealed the presence of a growth-deficient VNO in the naked mole-rat *Heterocephalus glaber*. The authors propose that such a degenerate VNO may be a sign of a diminished role for pheromone communication in this species. Smith et al.'s (2007) hypothesis is supported by a previous study from Faulkes and Abbott (1993) which demonstrated that reproductive suppression in *H. glaber* is solely behaviourally-induced. As discussed in Chapter 1, olfaction in Bathyergids, and in particular in naked mole-rats where it has been studied the most, is likely to be fundamental in a number of situations which would generally be mediated by pheromones in other species (Jarvis 1991, Reeve and Sherman 1991, Judd and Sherman 1996, O'Riain and Jarvis 1997). Thus, chemo-communication in naked mole-rats, and more generally in bathyergids, may be mediated by the MOE in a similar manner as that occurring in humans, since the VNOs in both species appear non-functional despite the behavioural support for pheromonal communication (Smith et al. 2007, Bhatnagar and Smith 2010). In this context it is interesting to have characterised a subpool of OR7 genes in Bathyergidae, since this gene family has been found to possess at least one gene in humans (OR7D4 as reported above, Wyart et al. 2007) that is directly involved in the binding of possible pheromone molecules. It is however important to note that the hypothesis of some ORs sensing pheromones does not exclude the possibility that VRs may be expressed in the MOEs of those species that have a degenerate VNO, possibly mediating chemo-communication (e.g. humans V1Rs are expressed in the MOE, Rodriguez et al. 2000).

2.4.3 The Bathyergid OR7 subgenome

In bats, evolutionary contemporaries of Bathyergids (Eocene - Teeling et al. 2005, Bennett & Faulkes 2000), analysis of OR family repertoires suggests a critical role for OR7 (grouped with OR families 1 and 3) in determining olfactory abilities (Hayden et al. 2010). This finding is based on the fact that ratios of functional genes:pseudogenes in OR1/3/7 explain most of the differentiation between bats and other mammals in a principal component analysis of all OR families (Hayden et al. 2010). Table 2.1

University of Cape Town

details the numbers of functional OR genes in the bat families analysed by Hayden et al. (2010), and includes the number of functional Bathyergid OR7 genes analysed in this study.

Table 2.1 Numbers of functional OR7 genes across bats
(Hayden et al. 2010)

Family	Number of functional OR7 genes
Phyllostomidae	48
Emballonuridae	25
Molossidae	11
Vespertilionidae	27
Mormoopidae	26
Rhinolophidae	2
Pteropodidae	63
Bathyergidae	51

Importantly, it emerges that the numbers of Bathyergid OR7 genes are comparable with those of the various bat families. This suggests that, although the subrepertoire characterised here is likely to underestimate global OR7 diversity in African mole-rats, it is large enough to be considered functionally important.

University of Cape Town

2.4.4 Functional variation across OR7 genes

All the OR phylogenies inferred in this study consistently recover the clustering of ORs according to four strongly supported gene lineages, corresponding to clades A-D (Figures 2.5 and 2.8). Importantly, functional ORs exhibit clade-specific motifs across the amino-acid sites involved in odorant-binding (Man et al. 2004, Katada et al. 2005), suggesting that ORs in each clade have different binding properties. Furthermore, it is interesting to note that, although the amino-acids involved in odorant-binding vary across clades, their chemical properties remain similar, with a remarkable prevalence of hydrophobic residues across the putative ligand-binding OR domain (Figure 2.6). This observation is consistent with Katada et al.'s (2005) hypothesis that the binding of odorant molecules into the odorant-binding pocket of ORs is mediated by hydrophobic interactions and van der Waals forces (Parsegian 2006). In their study, Katada et al. (2005) were able to predictably alter the binding properties of the mouse mOR-EG receptor (Kajiya et al. 2001) by inserting point-mutations at specific sites in the putative ligand-binding receptor domain. According

to their findings, the hydrophobic OR binding-pocket spanning TM3-6 constitutes a binding environment that is broad - i.e. able to recognise a range of odorants - but selective for the shape, size and length of odorant ligands. Therefore, the diversity observed across the odorant-binding domain of Bathyergid ORs in clades A-D is consistent with a scenario where different binding properties have been selected for within OR family 7, suggesting a functional importance of this gene family in African mole-rats.

Notably, the proportion of functional OR genes in each clade is variable, suggesting that different genetic lineages, within OR7, might have been important at different points in time during the evolution of the Bathyergid mole-rats. The evolutionary forces shaping OR diversity in the Bathyergidae OR clades is explored in the next chapter (Chapter 3), searching for historical signatures of selection across the African mole rat OR7 tree.

University of Cape Town

2.4.5 Trans-species polymorphisms

Trans-species polymorphism (TSP) i.e. the occurrence of similar alleles across related species (Klein et al. 1998) is a well documented phenomenon among genes of the Major Histocompatibility Complex (MHC), where most of the variation seen today is derived from ancestral species, as opposed to being generated *de novo* in each species following speciation (Figueroa et al. 1988, Klein 1987). Under this model of 'trans-species evolution', alleles that have a functional importance persist for long periods of time, resulting in a higher degree of relatedness between alleles across species, rather than within species (Takahata and Nei 1990). Consequently in a phylogenetic tree, alleles will tend to cluster according to allelic lineages rather than species, as seen for MHC genes in humans and chimpanzees (Mayer et al. 1992), fish (Garrigan and Hendrick 2001), birds (Richardson and Westerdahl 2003) and in the Bathyergidae (Kundu and Falukes 2007).

In this study, the occurrence of allelic variants of the same OR genes, as well as identical OR sequences across mole-rat species (as defined in section 2.2.6 of this chapter) are considered TSPs as per Klein et al.'s definition (1998), and suggest that a proportion of the genetic variability observed in Bathyergidae might be of ancient

origin. This finding is further supported by the fact that OR sequences do not cluster in a species-specific manner, but rather according to distinct genetic lineages corresponding to clades A-D (Figures 2.5, and 2.8), further suggesting the presence of ancient Bathyergid OR loci.

A similar trend has been reported by Zhang et al. (2007a) in a broad comparison of ORs and V1Rs between Mouse and Rat. The authors found that, in a phylogenetic tree, more than 99% of putatively functional ORs clustered into gene families with counterparts from the other species. In contrast, an analogous tree constructed with V1R sequences revealed the presence of species-specific V1R gene families. The authors ascribe these observed differences between ORs and V1Rs to the different evolutionary rates at which the two receptor types evolve. In their view, ORs evolve less rapidly than V1Rs because their principal role is to detect environmental odours, as opposed to V1Rs which detect pheromones and are therefore prone, by definition, to species-specific differentiation. However, we will see in the next chapter that this idea of 'slow evolution' as an explanation for ancient OR loci and TSPs is both simplistic and potentially inaccurate, since OR genes are found to evolve rapidly under a 'volcanic birth and death' model of evolution (Kambere and Lane 2007) which is governed by a complex range of mechanisms.

In conclusion, the isolation and characterisation of a subset of OR7 genes in Bathyergidae respresents the starting point to explore the genetic underpinnings of a well-developed olfaction, and raises a number of interesting evolutionary questions that are addressed in the following chapters.

University of Cape Town

Chapter 3: Evolutionary forces shaping Bathyergidae OR7 diversity

3.1 Introduction

3.1.1 Mechanisms of evolution of OR subgenomes

The size of OR repertoires varies widely across the genomes studied to date, as mentioned in Chapter 1 (see Figure 1.5 in Chapter 1 for details). In comparison with teleost fish, terrestrial vertebrates possess a 10-fold increase in the number of OR genes (Niimura and Nei 2005b); this increase is assumed to reflect the adaptation of the vertebrate olfactory system to the requirements of non-aquatic habitats (Glusman et al. 2001). OR subfamilies across tetrapods are functionally diverse, with lineage-specific patterns of expansion generating gene clusters of varying sizes in different species (Glusman et al. 2001, Young et al. 2002, Ben-Aire et al. 1994, Sullivan et al. 1996). It is hypothesised that the vertebrate OR superfamily evolves rapidly via a 'birth-and-death' model (Nei et al. 1997, Nei & Rooney 2005, Kambere and Lane 2007). The mechanisms underpinning this model of evolution follow a linear process of: a) gene duplication, b) adaptive selection, c) accumulation of OR polymorphisms, and d) OR pseudogenisation.

a) Gene duplication

The first step in the dynamic process of OR evolution consists of numerous gene duplication events, which are most probably mediated by unequal crossing over during meiosis, and as a result, generate tandem arrays of closely-related OR genes (Zhang et al. 2004, Glusman et al. 2001, Young et al. 2002). Segmental chromosome duplications may lead to a wide distribution of OR clusters, which are found on almost all chromosomes in the mouse and human genomes (Trask et al. 1998a, Rouquier et al. 1998, Sullivan et al. 1996). OR duplication events are likely to be facilitated by retrotransposon activity, since a high density of retrotransposons is found within OR gene clusters. These retrotransposon-dense regions indicate a history

University of Cape Town

of frequent DNA breaks during retrotransposition, and represent a fertile ground for repeat-mediated misalignments which generate unequal crossovers during meiosis (Glusman et al. 1996, Glusman et al. 2000b, Glusman et al. 2001). It has been suggested that gene conversion and recombination events may further contribute to the expansion of OR repertoires by creating 'mosaic' receptors and in doing so may bring about the 're-birth' of OR pseudogenes (Sharon et al. 1999). Conversely, gene conversion could also homogenise existing OR gene pools by substituting paralogous OR sequences with one another, resulting in decreased overall variability (Sharon et al. 1999). Therefore, gene conversion and recombination events appear to contribute only marginally, if at all, to the dynamic processes of OR gene expansion (Nei and Rooney 2005).

b) Adaptive selection

The number of functional OR genes of a species is thought to be proportional to the range of olfactants that can be detected and discriminated, thus reflecting a species odour recognition abilities (Niimura and Nei 2006, Kishida 2008). Given that odour discrimination can directly influence species' fitness (Munday et al. 2009, Dixson et al. 2010), it is reasonable to speculate that selective forces may act to enhance OR variability in species in which olfaction is fundamental.

A signal of positive (adaptive) Darwinian selection is generally defined by a significantly higher substitution rate of non-synonymous (dN, amino-acid replacing) to synonymous (dS, silent) mutations (Li 1997). The ratio of the two rates, dN/dS, denotes the magnitude and direction of selective pressure on a protein, with dN/dS =1, dN/dS <1 and dN/dS >1 indicating neutral evolution, purifying selection and positive selection respectively. Positive selection can further be subdivided into directional and balancing selection, the former favouring fixation of advantageous alleles, and the latter acting to maintain certain polymorphisms at optimal rates (Griffiths et al. 2000). The neutral theory of molecular evolution predicts that purifying selection is ubiquitous throughout the coding genome, whilst positive selection is destined to be rare (Kimura 1983). It has been argued that the increasing availability of sequenced DNA information has led to the misidentification of positive selection at genomic regions where no biological explanation supports this kind of selective pressure

University of Cape Town

(Hughes 2007). Only a strong *a priori* hypothesis should therefore justify the search for positive selection on any given protein (but see Zhai et al. 2012).

In his review paper, Hughes (2007) discusses how the most commonly used statistical methods are designed to detect a form of positive selection that favours repeated non-synonymous changes in a set of codons over long periods of time. This kind of selection usually occurs in genes involved in molecule–molecule recognition, typically involving a co-evolutionary process between e.g. a receptor and its ligand (Hughes 2007). Fulfilling the above prerequisites proposed by Hughes (2007), genes of the Major Histocompatibility Complex, MHC, known to evolve under a 'birth and death' model, are the ideal candidates for the study of adaptive evolution (Hughes and Nei 1989, Klein et al. 1993, Nei et al. 1997). MHC genes encode for cell-surface receptors which are involved in the recognition and binding of foreign peptides, thereby triggering an immune response via T-cells (Klein and Horejsi 1997). Intuitively, a parallel can be drawn between OR genes and MHC. Like the immune system, the olfactory system functions via interactions with the environment, thus evolving in concert with changes in the environment (Kambere and Lane 2007). It is therefore reasonable to hypothesise that, like MHC genes, positive selection would play a crucial role in OR gene evolution, driven by the necessity to recognise ecologically important odorants across different species and different habitats.

Indeed it has been demonstrated that a signal of positive selection characterises the portions of OR genes involved in ligand binding from fish to mouse and rat ORs, whilst the rest of the receptors' structure appears to evolve under purifying selection (Alioto and Ngai 2005, Emes et al. 2004, Kondo et al. 2002). This suggests that, after gene duplication events, positive Darwinian selection acts on the ligand-binding sites of duplicated OR genes to generate new binding properties, while strong purifying selective pressures maintain the overall functional structure of the receptors (Kambere and Lane 2007).

c) Accumulation of OR polymorphisms

Nucleotide polymorphisms, presumably generated by the same adaptive pressures that favour OR variation across species, are responsible for the high allelic variability of OR genes observed within species (Trask et al. 1998b). Interestingly, it has been suggested that allelic OR variation may account for observed differences in odour

University of Cape Town

perception among individuals in both humans (Wysocki and Beauchamp 1984, Gross-Isseroff et al. 1992) and mice (Griff and Reed 1995, Zhang and Firestein 2002). In humans, a comparison of 32 OR loci showed that Europeans and Ashkenazi Jews – referred to as 'Caucasians' - have a higher frequency of non-functional OR alleles when compared to African Pygmy populations (Gilad and Lancet 2003). A similar correlation of allelic OR functionality with ethnicity was found in a study of 51 human OR loci (Menashe et al. 2003), suggesting that modern human ORs have evolved dynamically in response to different environments and diets during recent human migrations. Similarly, a study on 16 OR genes across 20 breeds of dogs revealed the presence of breed-specific alleles (Tacher et al. 2005), which may correlate to the different olfactory abilities that have been artificially selected for in various working dog breeds e.g. scent hounds versus sight hounds.

Copy number polymorphisms represent yet another level of functional OR variation within populations (Trask et al. 1998b). Under this scenario, some individuals within a population possess 'extra' copies of OR genes due to recent gene duplication events. In human populations, for example, a study of 45 individuals from eight different ethnic groups revealed that a particular OR cluster, situated at the subtelomeric region of chromosome 19, was duplicated between 7-11 times in subtelomeres of other chromosomes (Trask et al. 1998b). The functionality of these copy number OR polymorphisms is, however, unknown.

d) OR pseudogenisation

In the context of 'birth and death' evolution, duplicated ORs that escape adaptive selection, progressively lose function and undergo pseudogenisation (Nei and Rooney 2005). Theoretically, OR pseudogenes should eventually become unidentifiable due to their accumulated mutations, or will be lost i.e. deleted from the genome since they are essentially neutral (Li et al. 1981, Gilad et al. 2003). Nevertheless, high proportions of OR pseudogenes have been retained in many vertebrate genomes (see Figure 1.5 in Chapter 1).

Traditionally, the ratio of OR genes:pseudogenes has been used as a measure of olfactory acuity, with OR pseudogenisation considered to be proportional to the decline of olfactory aptitude (Godfrey et al. 2004, Ache & Young 2005, Kishida 2008). For example, primates are thought to have accumulated a higher proportion of

University of Cape Town

OR pseudogenes, compared to other mammals, in response to the acquisition of full trichromatic vision, which is assumed to have triggered a regression in primate olfactory abilities (Gilad et al. 2004, but see Matsui et al. 2010). Similarly, the duck-billed platypus, which has a unique sense in its bill that combines electroreception and mechanoreception, enabling the species to forage with its eyes, ears and nostrils closed (Pettigrew 1999), has accumulated more OR pseudogenes in comparison with other mammals. This increased rate of OR pseudogenes relative to other species is intuitively linked to the platypus' allegedly limited reliance on olfaction for its most vital tasks (Nei et al. 2008). This pattern is also observed in toothed whales, in which high proportions of OR pseudogenes are associated with the evolution of echolocation as a response to full adaptation to a marine lifestyle (Freitag et al. 1998, Kishida et al. 2007, McGowen et al. 2008).

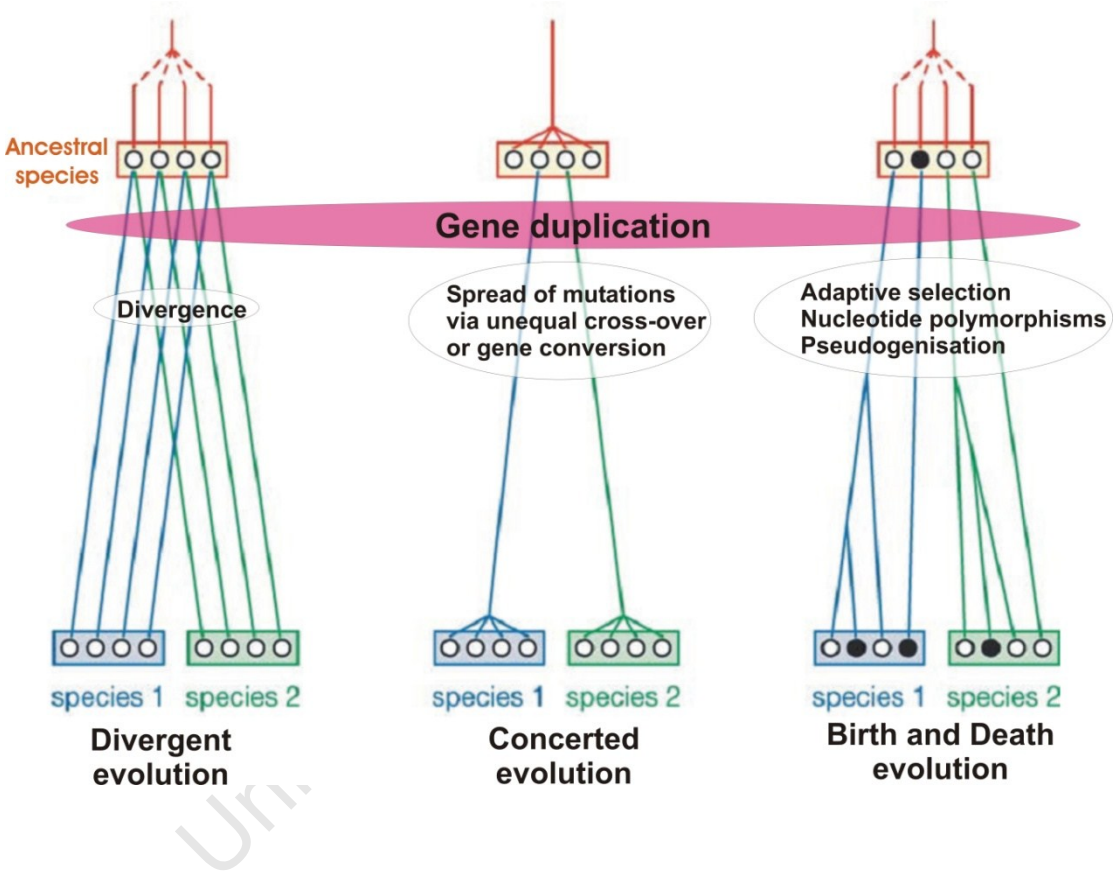
The evolution of OR functional genes and pseudogenes may however be more complex than suggested above. In a recent study, Zhang et al. (2007b) found a large proportion of human OR pseudogenes (67%) to be transcribed in the Main Olfactory Epithelium, MOE. Notably, a regulatory role in gene expression has been proposed for RNAs that are transcribed from pseudogenes (Balakirev and Ayala 2003, Duret et al. 2006, Zheng and Gerstein 2007). If transcribed OR pseudogenes have indeed such regulatory functions, it would explain why such large numbers of nearly-pristine OR pseudogenes are maintained over long periods of time across genomes. Furthermore, there is at least one example of an OR gene (the human OR17-210, Lai et al. 2008) previously classified as non-functional, that has been found to translate into an expressed olfactory receptor protein. This suggests that there may be more atypical functional genes scattered through what appears to be the vast sub-repertoire of OR pseudogenes (Nei et al. 2008).

Figure 3.1 illustrates three different evolutionary models known to occur in multigene families: divergent evolution, concerted evolution and 'birth and death' (described above). Divergent evolution is the most classic form of multigene family evolution, characterised by a gradual evolution of duplicated genes, leading to the acquisition of new functions (Nei and Rooney 2005). Concerted evolution, on the other hand, is described as "a form of multigene family evolution in which all the member genes are assumed to evolve as a unit in concert and a mutation occurring in a repeat spreads

University of Cape Town

through the entire member genes by repeated occurrence of unequal crossover or gene conversion" (Nei and Rooney 2005).

Figure 3.1 Models of evolution of multigene families (redrawn from Nei and Rooney 2005). Open circles represent functional genes and filled circles (black) represent pseudogenes.



3.1.2 Aims of this chapter

In this chapter, the different evolutionary forces that have shaped the Bathyergid OR7 subrepertoire are explored. Given the known dependence of African mole-rats on olfaction (Chapter 1), it is highly probable that positive selection has played a dominant role in generating functional variability across Bathyergidae OR genes. To test this hypothesis, a series of phylogenetic methods are used to identify signatures of selection across the ligand-binding domain of Bathyergidae ORs. In particular, the question of whether adaptive evolution has operated differentially on distinct clades of Bathyergidae genes is tested. Tests for positive selection using tree-based methods are applied to the Bathyergidae OR gene tree (Chapter 2, Figure 2.5); a branch-site test which allows dN/dS ratios to vary between branches is used to investigate signatures of adaptive evolution that may have acted on specific OR7 lineages over time**.** The results are interpreted within the framework of Nei's 'birth and death' model of evolution (Nei et al. 1997).

University of Cape Town

3.2 Methods

3.2.1 Signatures of selection on bathyergid OR clades

An average of substitution rates across the entire OR gene does not provide an accurate indication of positive selection, because olfactory receptors display a highly conserved overall structure with variability limited to a set of amino acid residues involved in the binding of odorant molecules (Gaillard et al. 2002, Katada et al. 2005). Therefore, tests for positive selection were performed at different codon sites in the dataset using the SELECTON server (available at http://selecton.tau.ac.il/ - Doron-Faigenboim et al. 2005, Stern et al. 2007). Estimates of the ratio of non-synonymous (dN) to synonymous (dS) substitutions are obtained for each codon, and significance is assessed via a likelihood ratio test (LRT). The LRT compares two nested models for each codon: a null model (M8a) which assumes no selection, and an alternative model (M8) which allows positive selection to occur. Three sets of bathyergid OR7 genes, corresponding to clades A, C and D, were analysed separately using codon-based multiple sequence alignment (MSA); pseudogenes were not included in the analysis. Clade B displayed only three putatively functional OR genes, and was therefore excluded from this analysis due to insufficient sample size. Across clade D, several codons with dN/dS ratios <1 are found, while none appear to have dN/dS > 1. A codon-based-Z test to test for purifying selection (overall average) was performed on clade D's functional ORs using the Nei-Gojobori method (Nei and Gojobori 1986) implemented in MEGA5 (Tamura et al. 2011) with 1000 bootstrap replicates to determine significance of purifying selection across this clade.

3.2.2 Signatures of selection across the bathyergid OR gene-tree

To investigate whether positive selection may have acted along specific Bathyergid OR lineages, a branch-site test (test 2 in Zhang *et al.* 2005) was carried out in CodeML, based on the phylogenetic tree of African mole-rat OR genes (Figure 2.5, Chapter 2). CodeML is a program within the Phylogenetic Analyses by Maximum Likelihood (PAML) package (Yang 1997, Yang 2007) that estimates the dN/dS ratio (ω) on codon (nucleotide) alignments, based on gene tree topology. In the branch-site test of

University of Cape Town

positive selection, a branch of interest, called a 'foreground branch', is isolated from the other 'background' branches. The value of ω is fixed among codon sites in the background branches, while it is allowed to vary in the foreground branch. Two nested models, null and alternative, are computed and compared using a LRT. In the null model, codons along all branches are either under purifying selection ($\omega < 1$) or under neutral evolution ($\omega = 1$), and the foreground branch may have different proportions of sites under neutral selection than the background branches (i.e. relaxed purifying selection). In the alternative model, some sites on the foreground branch may be under positive selection ($\omega > 1$). For this analysis, a maximum likelihood tree (GTR, 1000 bootstrap) (Felsenstein 1981, 1985, Tavaré 1986) was constructed in MEGA v5 (Tamura et al. 2011) using the 119 OR genes identified in this study; following Yang (2009), stop codons and alignment gaps were excluded from the alignment in Bioedit v7.0.8.0 (Hall 1999). The resulting tree maintains the same tree topology as the original bathyergid OR tree constructed in Chapter 2, based on full length nucleotide alignment (Figure 2.5 and Appendix III.2). In the branch-site test of positive selection, each branch of the OR gene tree was labelled in turn as foreground; a LRT was performed on all pairs of nested models and compared to a χ^2 distribution to determine significance. Furthermore, the Q-value - a measure of the false discovery rate due to multiple testing - was calculated for each branch using the 'Q-value' software available at http://genomics.princeton.edu (Storey 2002, 2003, Storey et al. 2004). When the LRT was significant (p-value< 0.05) and the false discovery rate was low (i.e. less than 15%), the posterior probability of sites being under positive selection ($dN/dS > 1$) was calculated using the Bayes Empirical Bayes (BEB) method (Nielsen and Yang 1998, Yang et al. 2005b) implemented in CodeML.

University of Cape Town

3.2.3 The ratio of functional OR7 genes:pseudogenes

The ratios of functional OR7 genes: pseudogenes across clades A-D, together with the number of putatively functional genes and pseudogenes in each clade will be considered in this chapter. Both are expected to be consistent with the different scenarios of evolution that emerge from the analysis reported above, and will be used to contextualise the results in an evolutionary framework.

3.3 Results

3.3.1 Signatures of selection on African mole-rat OR clades

Tests for differential positive selection across bathyergid OR clades revealed a number of interesting patterns. Likelihood ratio tests for positive selection were performed on the functional OR7 genes from clades A, C and D, while clade B was excluded from this analysis due to insufficient sample size.

A summary of the LRT results is reported in Table 3.3 (a detailed table of results can be viewed in Table 1 – Supplementary Material on CD).

Table 3.3 Summary of likelihood ratio tests - LRTs - across OR clades (total of 212 codons analysed)

Clade	Number of positively selected sites (dN/dS>1)	LRT
A	24	Significant
C	34	NS*
D	0	NS*

*NS= non significant

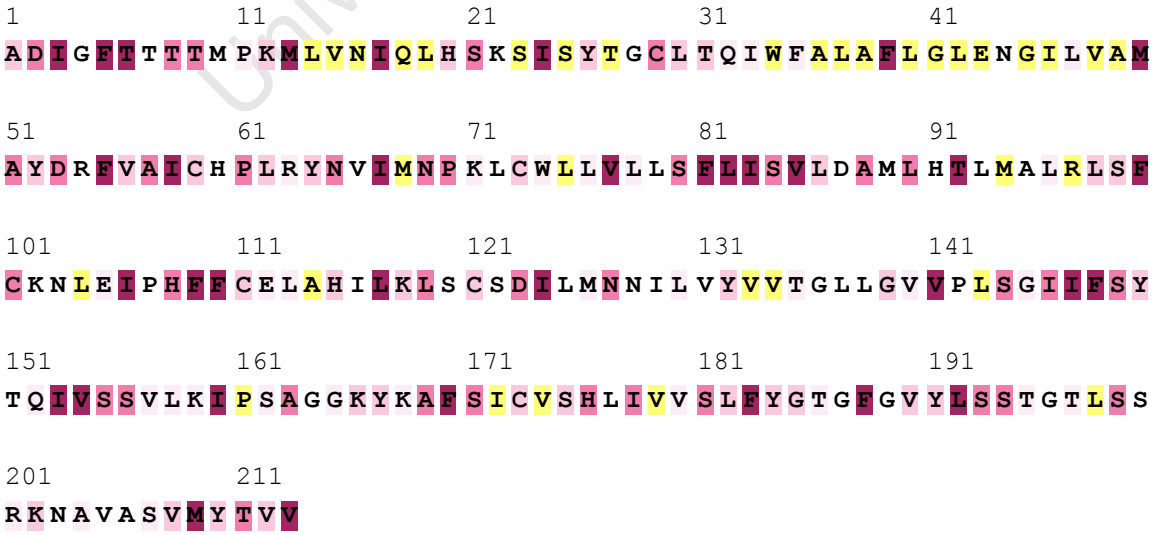
A graphical comparison of positive selection across different codon sites in clades A, C and D, based on dN/dS ratios, is shown in Figure 3.3. At first glance, purifying selection represents the dominant visual pattern for all clades (pink and purple residues); for this reason, dramatic emphasis is attributed to positively selected sites in clades A and C (yellow and orange), whilst no codons across clade D are consistent with adaptive selection.

University of Cape Town

Figure 3.3 Selection across clades A, C and D The first sequence of each clade's nucleotide alignment is displayed, as a consensus. Amino-acid residues are coloured according to dN/dS ratios at codon sites; dN/dS<1 pink and purple – purifying selection, dN/dS>1 yellow and orange – positive selection. The putative odorant-binding site of ORs, TM3-6, corresponds to the stretch between amino-acid positions 25-198 (Man et al. 2004, Katada et al. 2005).

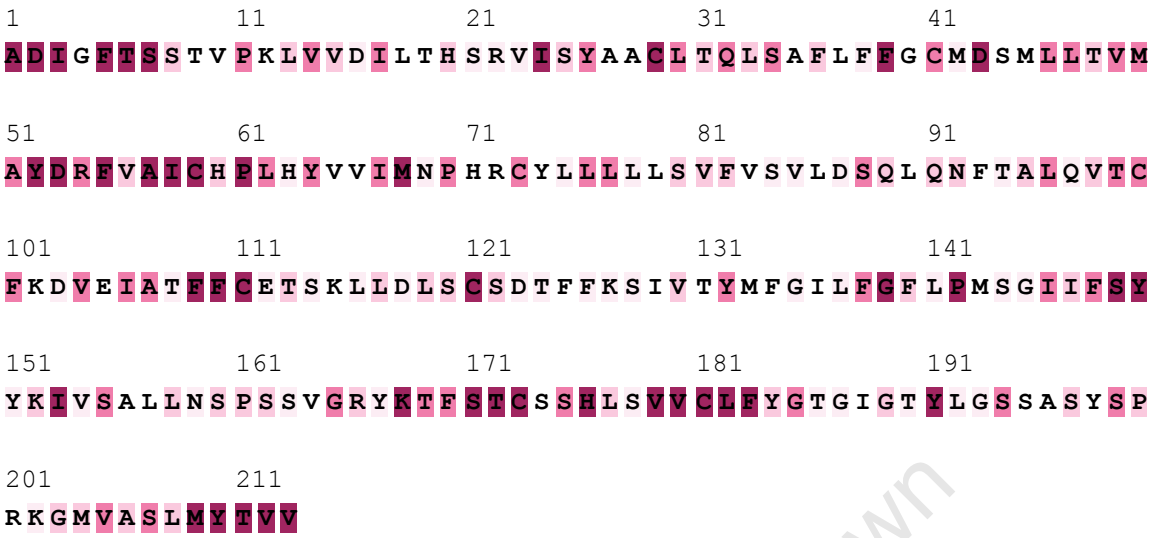
Clade A

Clade C



University of Cape Town

Clade D



Legend:

The selection scale:

1 2 3 4 5 6 7

Positive selection Purifying selection

University of Cape Town

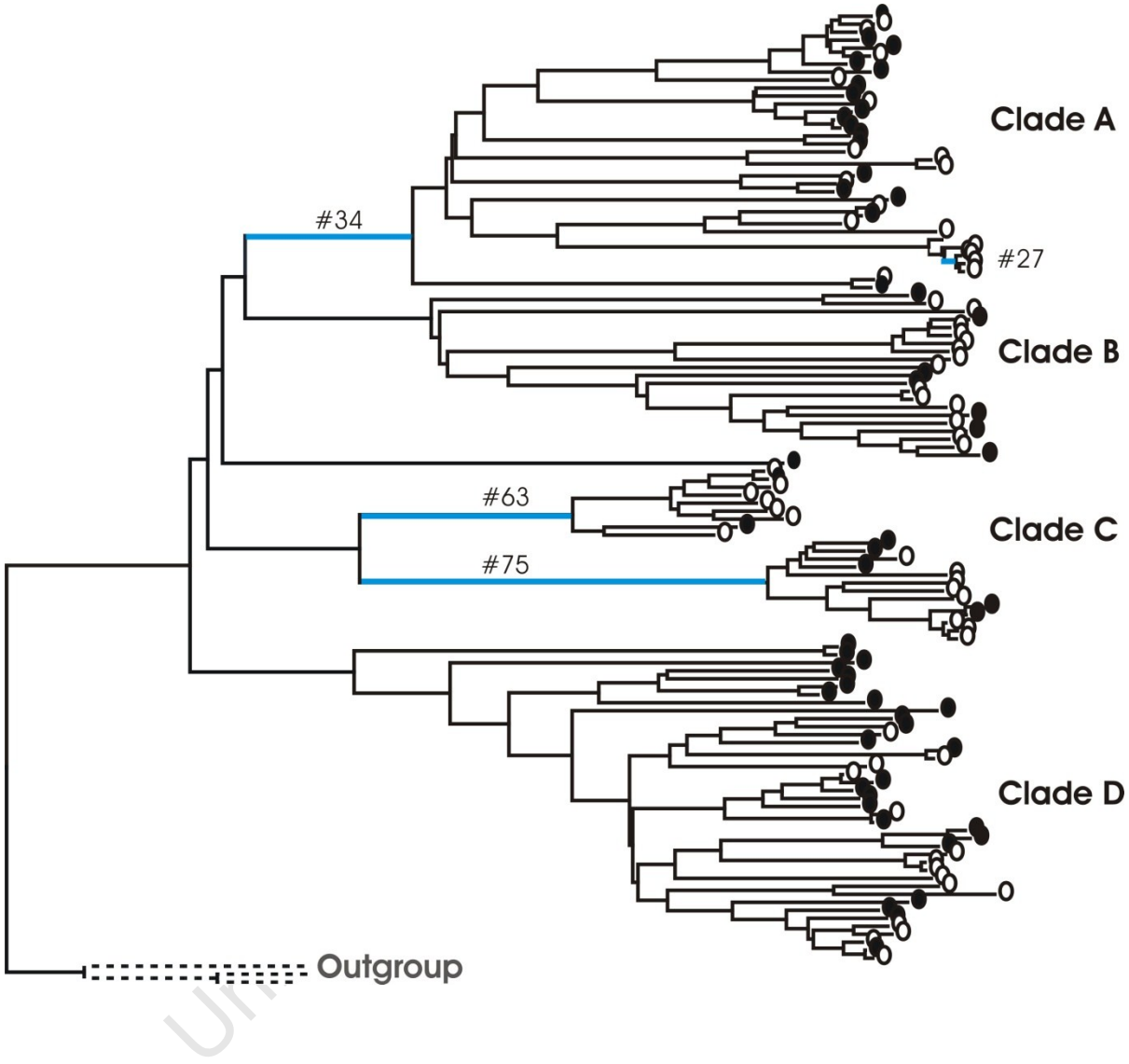
According to the dN/dS ratios of codon sites, 24 codons appear under positive selection in clade A, and the LRT supports a significant signal of positive selection across this clade (Table 3.3). Although 34 codon sites were identified as evolving under positive selection in clade C based on the dN/dS ratios (Table 3.3, Figure 3.3), the LRT failed to identify an unambiguous signature of positive selection in this clade. On the other hand, the majority of sites along clade D are characterised by dN/dS ratios <1 and no sites under positive selection are evident (Table 1 – Supplementary Material), suggesting a prevalence of purifying selective forces acting along the OR genes of this clade. Clade D was thus tested separately for purifying selection via a codon-based Z test (1000 bootstrap, Nei-Gojobori method; Nei and Gojobori 1986) and a strong signal of purifying selection was confirmed for this clade ($p < 0.0001$).

3.3.2 Lineages under positive selection along the OR gene-tree

To identify episodic events of adaptive evolution across specific Bathyergidae OR lineages, a branch-site test of positive selection, that allows dN/dS to vary across lineages in a phylogenetic tree (Yang and Nielsen 2002, Zhang et al. 2005), was performed across all branches of the African mole-rat OR tree (Figure 2.5). The full list of results for the branch-site test is reported in Appendix I.7. Six branches in the tree support a signal of positive selection in the corresponding lineages ($p<0.05$). When Q-values are taken into account, however, positive selection can only be inferred unequivocally for four branches, with a false discovery rate <14% (these correspond to branches 75, 34, 27 and 63 in Appendix I.7). These four lineages under positive selection in the Bathyergidae OR gene tree are indicated in Figure 3.4.

University of Cape Town

Figure 3.4 Positively selected lineages in the Bathyergidae OR gene tree Positively selected lineages, according to branch-sites analysis, are coloured in blue; # branch numbers correspond to those assigned by CodeML (Appendix I.8); filled circles (black) represent functional OR genes, open circles represent OR pseudogenes.



Results from a BEB (Nielsen and Yang 1998, Yang et al. 2005b) analysis to identify which amino-acid sites are evolving under adaptive evolution revealed that the number and location of positively selected sites vary among these four lineages (reported in Table 3.4).

Table 3.4 Positively selected residues in Bathyergidae OR lineages Results from the Bayes Empirical Bayes analysis are reported under the BEB column and indicate the probability of the corresponding amino-acid residues being under positive selection. Branch numbers match labelled branches in Figure 3.4. ♦ Amino-acid positions and location domains were assigned based on the molecular model developed in Chapter 2 (Figure 2.4).

Clade	Branch	Amino-acid position ♦	BEB	Location domain ♦
A	34	30	0.744	TM3
		99	0.754	EC2
		102	0.809	EC2
		158	0.778	IC3
		162	0.912	IC3
		186	0.983	TM6
		202	0.818	EC3
		205	0.657	TM7
A	27	7	0.993	TM2
C	63	20	0.595	EC1
		177	0.542	TM6
		179	0.843	TM6
C	75	7	0.649	TM2
		21	0.882	EC1
		37	0.942	TM3
		40	0.86	TM3
		41	0.807	TM3
		46	0.829	TM3
		47	0.805	TM3
		48	0.775	TM3
		49	0.851	TM3

3.3.3 Ratios of functional OR7 genes: pseudogenes across clades

The numbers and ratios of functional OR7 genes and pseudogenes across clades A-D are reported in Table 3.5; these are expected to be consistent with the different evolutionary forces found across bathyergid OR7 genes, and will be considered in the discussion to follow.

Table 3.5 Numbers of functional ORs and pseudogenes in clades A-D

Clade	Functional ORs	Pseudo-ORs	% Functional ORs	% Pseudo-ORs
A	17	17	50	50
B	3	18	14	86
C	5	17	23	77
D	25	15	63	37

University of Cape Town

3.4 Discussion

Given the direct association of functional OR diversity with olfactory ability (Niimura and Nei 2006, Kishida 2008), and the known importance of olfaction for bathyergids (Chapter 1), positive Darwinian selection is predicted to play a fundamental role in driving observed variability at OR loci within the African mole-rats. The hypothesis of adaptive evolution acting on mole-rat OR genes follows Hughes (2007), whereby adaptive evolution is tested within a gene family involved in molecule-molecule recognition, and where repeated non-synonymous changes in a set of codons may be favoured over long periods of time. Such conditions arise from the constant need of organisms to identify and respond to olfactants in their social and physical environments, providing support for an *a priori* hypothesis of positive selection on those OR genes that account for niche-specific odorant-recognition. Simultaneously, olfaction also plays an essential role in the biology of species by enabling the detection of fundamental odorants which may remain unaltered through space and time (e.g. the smell of death or rotten foods). Thus in this context, an expectation of purifying selection is reasonable.

A series of classical tests for the detection of non-neutral molecular evolution were performed across different portions of the bathyergid OR7 gene family and reveal that disparate selective forces have influenced its evolution, including both positive selection (clades A and C), and purifying selection (clade D). Signatures of adaptive evolution among OR lineages were identified via branch-site analyses (Yang and Nielsen 2002, Zhang et al. 2005) across the bathyergid OR gene tree. Interestingly, all four episodes of adaptive evolution identified in the bathyergid gene tree are assigned to branches which lead to the clades that display positively selected codon sites (dN/dS>1), namely clades A and C (Figure 3.4). Conversely, purifying selection represents the dominant force in clade D, where most of the trans-species polymorphisms were identified (Chapter 2). The biological significance of such diverse evolutionary forces shaping distinct sets of closely-related OR genes is discussed in the context of 'volcanic birth and death' evolution that characterises the OR multigene family (Kambere and Lane 2007).

University of Cape Town

3.4.1 Signatures of selection on bathyergid OR7 genes

In chapter 2, functional genes in clades A-D (Figure 2.5) were found to share clade-specific amino-acid motifs across the ligand-binding site of ORs, suggesting that each clade may have different binding properties. In order to unravel the evolutionary mechanisms responsible for this scenario, an analysis of the signatures of selection revealed on these clades is discussed, by considering the ratios of functional OR genes: pseudogenes and the numbers of functional OR genes isolated in each clade.

- **Clade A**

In clade A, an explicit signal of adaptive selection was detected, with the majority of sites under positive selection (dN/dS>1) concentrated in the region between TM3-TM6 (Figure 3.3), i.e. across the odorant-binding site region (Gaillard et al. 2002, Katada et al. 2005). Importantly, the ratio of functional OR genes to pseudogenes in this clade - 50 %:50 % - is relatively higher than that of other clades, as is the number of functional ORs (Table 3.5). This suggests that selection for functional variability among ancestral bathyergid OR genes is likely to have occurred within this clade, revealing an important role for these genes in mole-rat olfaction.

When episodic events of positive selection were examined via a branch-site test along the gene-tree, a strong signal of adaptive evolution was found on the branch at the root of clade A (branch #34, Figure 3.4). Along this branch, six out of the 8 amino-acid sites that were identified as evolving under positive selection lie in TM3-6 region (Table 3.4), suggesting that selective forces have acted predominantly on the odorant-binding region, presumably to generate novel binding properties (Man et al. 2004, Katada et al. 2005). The second branch that carries a signal of positive selection in this clade is # 27 (Figure 3.4), which leads to a subset of *H. glaber* OR pseudogenes (Appendix III.2). In this case, only one positively selected codon was identified (Table 3.4), within TM2, which is a relatively conserved portion of OR genes (Katada et al. 2005). This, together with the fact that only pseudogenes are present in the sub-clade that generates from branch #27, suggests that the signal of positive selection detected by the branch-site test might in fact be a consequence of pseudogenisation, rather than adaptive evolution.

University of Cape Town

- **Clade B**

In addition to a small number of putatively functional ORs, Clade B has the highest proportion of pseudogenes (86%, Table 3.5), suggesting that ORs within this clade may be secondary for bathyergid olfaction and prone to pseudogenisation. Along with this view, the branch-site test of positive selection failed to detect any episodic events of positive selection across this clade.

- **Clade C**

The proportion of functional OR genes in clade C is lower than that of clade A (23%), as is the number of functional OR genes (Table 3.5). Although the LRT fails to identify an unambiguous signal of positive selection on this clade, several codon sites are characterised by dN/dS values >1 along the ligand-binding site of ORs (Figure 3.3). Interestingly, the branch-site test of positive selection identifies two ancestral branches across clade C - # 63 and # 75 – with a strong signal of adaptive selection at numerous amino-acid locations, within the odorant-binding domain of ORs (mostly in TM6 for branch #63 and TM3 for branch #75, Table 3.4). These results are consistent with an ancestral signal of positive selection on clade C ORs, perhaps indicating a phase when new OR functionalities were acquired along this gene lineage.

- **Clade D**

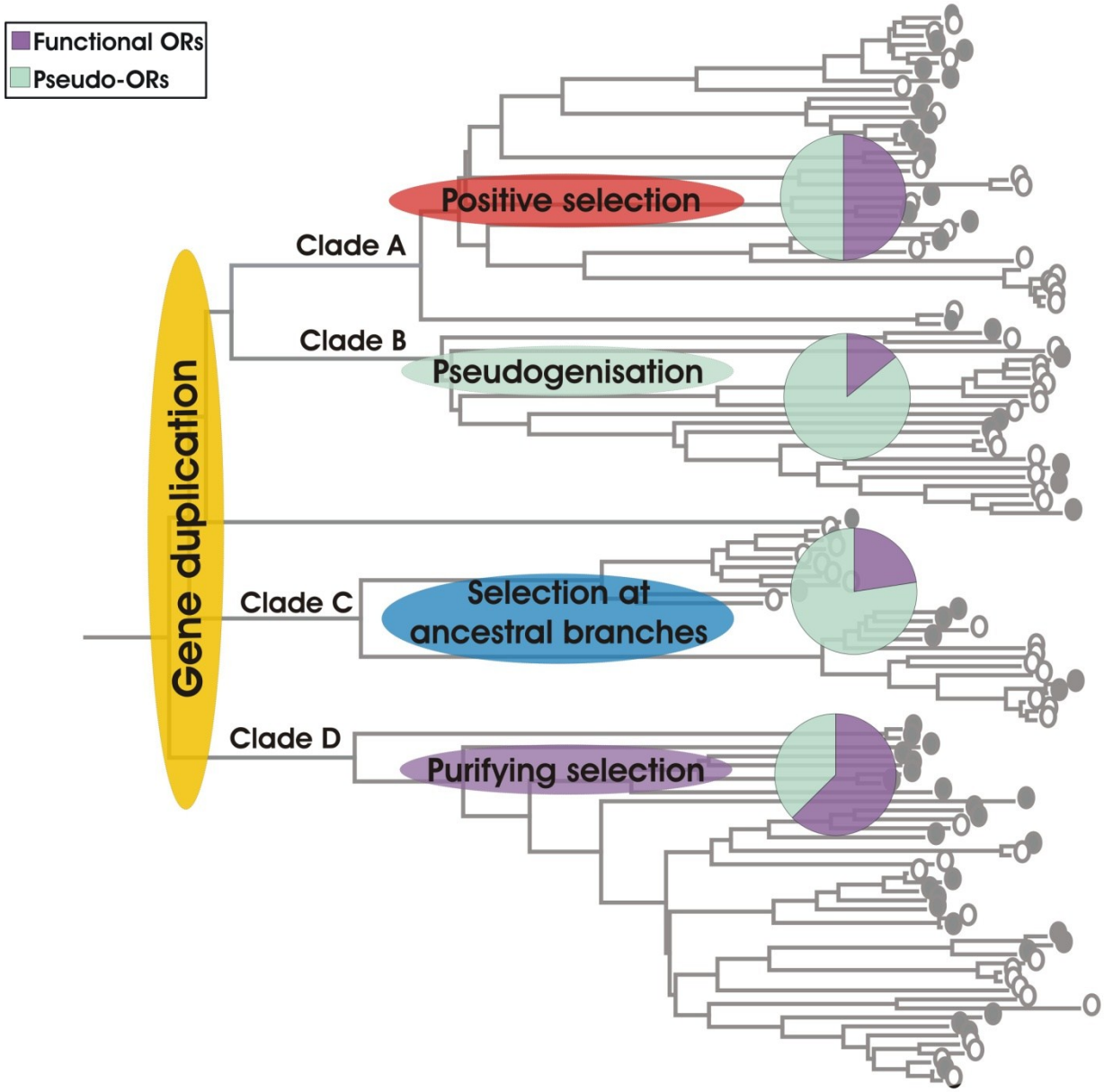
Along Clade D, finally, it appears that strong purifying selection has maintained unaltered over a long period of evolutionary time this subpool of OR genes, from the divergence of the major mole-rat genera *Bathyergus*, *Georychus*, *Cryptomys* and *Fukomys* onwards (17-15 MYA, Ingram et al. 2004, Chapter 1 - Figure 1.7) and throughout the phylogeny (Appendix I.5), with no apparent signal of adaptive evolution. Interestingly, the highest proportion of functional ORs (62.5%), as well as the greatest number of putatively functional genes, are found in this clade (Table 3.5). These results suggest a scenario where odorant chemicals that carry fundamental information for Bathyergidae fitness are recognised by clade D ORs, and are therefore actively maintained unchanged in time.

University of Cape Town

The different evolutionary forces acting on Bathyergid OR7 genes across clades A-D are summarised in Figure 3.5. Given the framework of 'birth and death' evolution characteristic of OR genes (Kambere and Lane 2007), it is probable that a number of gene duplication events characterised the genesis of the mole-rat OR7 family. Sets of similar OR genes, generated by gene duplications, likely differentiated in function according to the selective forces described above, driven by specific biological constraints.

University of Cape Town

Figure 3.5 Summary of the selective forces acting on Bathyergidae OR7 gene family. Gene duplication events likely occurred at the origin of the mole-rat OR7 gene family ('birth and death' evolution, Kambere and Lane 2007). Pies represent the proportions of functional OR genes (purple) and OR pseudogenes (light green) in clades A-D; the tree on the background is the bathyergid OR gene tree used for Figures 2.5 and 3.4.



3.4.2 The role of positive selection on bathyergid OR genes

In theory, positive selection is expected to maintain part of the observed functional variability at OR loci (Kambere and Lane 2007). In particular, positive selection is predicted to act on the ligand-binding region of ORs, while the overall receptor structure typical of GPCRs is expected to be maintained by purifying selection (Emes et al. 2004, Alioto and Ngai 2005, Kambere and Lane 2007). Consistent with the above, our results for clade A are similar to those previously reported for other vertebrate species (Alioto and Ngai 2005, Emes et al. 2004, Kondo et al. 2002), with positive selection acting predominantly on the ligand-binding domain of mole-rat OR7 genes. Similarly, along clade C, the two ancestral branches that carry a significant signal of adaptive evolution across the receptors' ligand-binding region are likely indicative of a historic spurt of selection on these OR loci.

Such signals of positive selection detected along the bathyergid OR gene tree suggest that functional variation on these loci was generated in an effort to enhance the range of odorants recognised, and/or to optimise the recognition of crucial odorants. This implies that the detection of such odorant molecules may be directly related to fitness. Under this scenario, adaptive evolution is likely an indicator of intra-specific competition for olfactorily-mediated resources, as hypothesised by Emes et al. (2004). In Emes et al.'s view (2004), OR gene duplication and sequence diversification via positive selection are driven by competition between individuals e.g. for food or predator avoidance. Unfortunately, information on specific ORs and their odorant ligands is generally scarce and it is therefore difficult to establish an unequivocal link between OR diversity at specific loci and fitness (Zarzo 2007). A theoretical association between OR variation and fitness is however indisputable, due to the need of ORs to recognise odorants from an ever-changing environment in a way that is comparable to the MHC-pathogen co-evolution, as mentioned in this chapter's introduction (Kambere and Lane 2007). Whether it is the environment or bathyergid species-specific life history traits that act as drivers for enhancing functional OR variation will be the subject of the next chapter.

University of Cape Town

3.4.3 Purifying selection and Trans-Species Polymorphisms (TSPs)

The majority of functional TSPs i.e. identical alleles of a same OR gene that are present across mole-rat species (introduced in Chapter 2, Klein et al. 1998), found at bathyergid OR7 loci occur within clade D (70%, Appendix I.5). According to the results reported above, clade D has evolved under strong purifying selection. Importantly, this clade has the highest number and proportion of functional ORs, suggesting that TSPs across this clade may represent allelic variants that maintain precise binding properties among bathyergid species, enabling them to detect fundamental olfactants.

The occurrence of TSP and ancient OR loci has previously been reported in a comparative study of whole Mouse and Rat OR subgenomes (Zhang et al. 2007a). According to the authors, the presence of conserved OR loci across species could be explained by 'slow OR evolution', i.e. the genes have evolved neutrally such that conserved loci are a consequence of relatively recent divergence between the species analysed; according to a molecular phylogeny based on multiple genes, *Mus* and *Rattus* diverged ~23 MYA (Adkins et al. 2001). In this study, the data suggests that purifying selection has actively maintained conserved OR loci as well as TSPs for a relatively shorter period of time. Indeed, functional conserved OR loci and TSPs are present from the divergence of the major mole-rat genera *Bathyergus*, *Georychus*, *Cryptomys* and *Fukomys* onwards (17-15 MYA, Ingram et al. 2004, Chapter 1 - Figure 1.7), and throughout the phylogeny (Appendix I.5).

Under neutral expectations, alleles are expected to become species-specific when species have diverged for more than 4Ne generations (where Ne is the effective population size; Kimura and Ohta 1969). Information on bathyergid population sizes is not available due to the difficulties in estimating the numbers of these subterranean mammals (Bennett and Faulkes 2000). It is however known that species are characterised by a relatively rapid turnover of generations, reaching sexual maturity, on average, within their first year of life (estimated based on the age at which individuals attain their adult weight - with the exception of *H.glaber*) and producing between 1-4 litters per year, depending on the species (Bennett and Faulkes 2000). Thus, if we take the minimum estimate of 1 generation per year, bathyergid population sizes would have to be of the range of 4 million individuals - Ne = ~16

University of Cape Town

million (Ingram et al. 2004) / 4 - to explain the persistence of TSPs under neutral evolution. Populations of this magnitude are clearly unrealistic.

It therefore seems reasonable to speculate that TSPs at OR loci in bathyergid mole-rats are most likely the result of purifying selection, suggesting that such polymorphisms are essential for the family and are likely to confer significant fitness benefits. Similarly, one could confidently envisage that at least some of the TSPs identified by Zhang et al. (2007a) across Mouse and Rat ORs are likely to be a result of selective pressures acting to maintain specific capacities of odour recognition among Muridae, rather than a by-product of neutral evolution, especially given the divergence time between the two species (Adkins et al. 2001).

3.4.4 Evolutionary constraints on ectopically expressed ORs

An alternative hypothesis that may explain the occurrence of purifying selection on particular OR genes comes from a recent study by De la Cruz et al. (2009).

Several genes predicted to be ORs are expressed in mammalian non-olfactory tissues (Parmentier et al. 1992, Vanderhaeghen et al. 1997, Branscomb et al. 2000, Feldmesser et al. 2006), suggesting the possibility that a subset of ORs may carry out additional non-olfactory functions when ectopically expressed. With the exception of a proposed role in chemotaxis for OR genes expressed in sperm (Spehr et al. 2003), the hypothesis of non-olfactory functions for ectopically expressed ORs is largely unsupported. Given the scarcity of experimental techniques allowing for the functional characterisation of OR proteins, De la Cruz et al. (2009) propose an evolutionary approach to address the question of whether ectopically expressed ORs perform alternative functions.

De la Cruz's (2009) evolutionary analysis relies on the hypothesis that functionally important traits evolve under evolutionary constraint. In this view, OR genes that may have alternative functions are expected to have similar patterns of ectopic expression in closely related species. Furthermore, because of the evolutionary constraint associated with the additional functions, orthologous OR genes that have conserved ectopic expression patterns across species are predicted to evolve at a slower pace compared to ORs that are exclusively expressed in the MOE (Duret and Mouchiroud 2000, Winter et al. 2004, Khaitovich et al. 2005).

University of Cape Town

By comparing expression data of human (Zhang et al. 2007b) and chimpanzee ORs from four non-olfactory tissues (liver, testis, lung and heart), De la Cruz et al. (2009) identified a subset of ectopically expressed ORs that display conserved expression patterns across the two species. All OR genes identified were expressed in both the main olfactory epithelium and one of the non-olfactory tissues analysed. To measure the different rates of protein evolution, De la Cruz et al. (2009) used dN/dS ratios, and found lower ratios i.e. slower rates of evolutionary change in OR genes with conserved patterns of ectopic expression, when compared to ORs that were expressed exclusively in the MOE. Thus, the authors conclude that the subset of ectopically expressed OR genes analysed have evolved under stronger evolutionary constraint than their exclusively olfactory counterparts. Even though this finding does not provide direct information on the function of ectopically expressed ORs, it supports the hypothesis that some OR genes may carry out additional functions in non-olfactory tissues.

In De la Cruz's (2009) study, a strong association is apparent between the intensity of purifying selection and the alleged 'atypical' functions of OR genes that are expressed in non-olfactory tissues. Interestingly, dN/dS ratios for mole-rat OR genes across clade D (median dN/dS = 0.51) are closer to those found for ectopically expressed ORs (median dN/dS = 0.38), than those of exclusively olfactory ORs (median dN/dS = 0.69, De la Cruz et al. 2009). This consideration is certainly inconclusive, because the bathyergid OR7 repertoire analysed here is not directly comparable to the OR repertoires from De la Cruz et al.'s study (2009). Nevertheless, the prospect that clade D ORs may be under stronger evolutionary constraint when compared to other mole-rat ORs due to a possibly non-olfactory role cannot be excluded. To further test this idea one would need to verify if these specific genes are transcribed in non-olfactory tissues. This could be achieved by extracting mRNA from the target tissues in the different mole-rat species, and using either a PCR-based approach like the one in this study, or preferably a whole-transcriptome sequencing approach followed by bioinformatic analysis to identify putative OR genes.

3.4.5 A new method for characterising OR subfamilies

The subpool of Bathyergidae OR genes characterised in this study belongs to a single family of ORs, namely OR7 (discussed in Chapter 2). In an attempt to carry out a

University of Cape Town

fine-scale classification of the complex OR gene superfamily, a number of recent studies subdivide OR families into 'subfamilies' based on 'pattern' i.e. setting sequence similarity cut-offs (generally 60%, Glusman et al. 2000a, Godfrey et al 2004). If we look at the average pairwise distances (based on the number of nucleotide differences) between functional ORs from clades A-D (Appendix I.6), between 62-68% of the sequence similarity occurs across clades. Therefore, if we were to classify Bathyergidae ORs into subfamilies according to 'pattern', with the generally accepted cut-off limit of 60% (Glusman et al. 2000a), the observed clade structure would not reflect subfamily structure, because all ORs would fall into a single subfamily. Nevertheless, the results presented here suggest that the clustering of ORs into clades that have evolved under disparate selective forces potentially reflects their underlying biological significance. Despite the high percentage of between-clades sequence similarity, there appears to be a strong functional association between genes belonging to each clade, making it tempting to speculate that, from a functional viewpoint, each clade represents a distinct OR7 subfamily. These findings raise the question of the appropriateness of a practice that has been very common in large-scale OR studies, of classifying OR genes into subfamilies based solely on sequence similarity criteria. I therefore propose that an investigation of the evolutionary mechanisms that shape OR genetic diversity across clades can be used as an additional, novel and perhaps more accurate approach to OR gene classification, based on 'process' rather than 'pattern' alone.

In summary, this analysis represents the first attempt to investigate the intricate mechanisms shaping the evolution of the African mole-rat olfactory repertoire. Within the OR7 family, mole-rat genes have been subject to a spectrum of evolutionary forces. In addition to the classic aspects of 'birth and death' evolution (Nei and Rooney 2005), an important role for purifying selection emerges in the gene family. The 'clade structure' observed in the Bathyergidae OR gene tree appears to support a 'subfamily structure' based on OR functional properties, and reflects the broad range of odorant ligands that mole-rat OR7 genes can recognise. Interestingly, these findings challenge the commonly accepted theory that closely related ORs share functional properties (Malnic et al. 2004). The proposed methodology of OR classification based on 'process' rather than 'pattern' promises to be generally applicable to OR analysis, offering a genuine tool for functional OR classification.

University of Cape Town

Chapter 4: The role of life-history traits and environment in shaping Bathyergidae OR diversity

4.1 Introduction

This study provides evidence that bathyergid OR genes have evolved under strong selection, supporting both the well-developed olfaction known to characterise the family and the expectations of natural selection shaping and maintaining functionality across vertebrate OR genes (Kambere and Lane 2007). In this chapter two different agents of selection, as well as the relative contributions of each agent, are explored within a modern bioinformatic framework. The approach considers both life history traits, specifically sociality, and environmental niche specialisation as putative, non-mutually-exclusive, drivers of natural selection acting on OR genes.

University of Cape Town

4.1.1 Sociality and Olfaction

The vertebrate olfactory system plays a fundamental role in both processing social odours and in the effects of learning in a social context (Sanchez-Andrade and Kendrick 2009). Social olfactory information detected by the main olfactory system is processed in specialised brain structures that control social behaviour, recognition, attraction and bonding (Sanchez-Andrade and Kendrick 2009). In line with this, an increasing number of social behaviours are found to be governed by the main olfactory epithelium (MOE) - thus potentially detected by ORs – in social mammalian species. For example, in sheep herds, ewes are able to recognise their lambs via a selective olfactory memory, which is imprinted in a sensitive period of 2-4 hours after birth (Kendrick et al. 1992), and results in ewes directing their parenting efforts exclusively towards their own offspring. This critical recognition is dependent on the main olfactory system (MOS): chemically interfering with the MOE - via zinc sulphate-induced anosmia (the inability to perceive odours) - affects the selectivity of the ewe's maternal behaviour, however sectioning of the vomeronasal organ (VNO) has no effect on the behaviour (Levy et al. 1995). In social rodents such as mice and

rats, individual recognition of conspecifics has been shown to occur via a mechanism of short- and long-term olfactory memory (Bluthe and Dantzer 1990, Axelson et al. 1999, Kogan et al. 2000), where individual 'identity' is specified by olfactory cues from urine, skin secretions, reproductive tract or specialised scent glands (Mykytowy and Goodrich 1974, Natynczuk and Macdonald 1994, Stopka et al. 2007). The recognition of such olfactory cues is conveyed by the MOS, since both lesions in the main olfactory bulb and chemically-induced anosmia interfere with social recognition (Dantzer et al. 1990, Popik et al. 1991, Bluthe and Dantzer 1992, Matochick 1998). In this context, ORs take part in the essential role of conveying vital social information. Because of the complex olfactory-mediated social behaviours that ORs are responsible for, it is thus plausible that sociality may be a driver of accelerated evolution at OR genes.

The most peculiar feature of the bathyergid mole-rats is that, as a mammalian family, they display such a broad range of social systems, from the strictly solitary systems of the genera *Bathyergus*, *Georychus* and *Heliophobius*, to the cooperative breeding and remarkable housekeeping of *Cryptomys, Fukomys* and *Heterocephalus*. Intuitively, one could envisage a scenario where olfaction plays diverse roles in bathyergid societies according to their social structure. Thus, it is plausible to propose that solitary species may be expected to have a more developed sense of smell in response to the challenges of finding food or mates when living a solitary lifestyle. On the other hand, a subterranean social lifestyle may require an acute sense of smell for dealing with situations that do not generally characterise solitary living, such as the need for recognising colony members and their relative social status, recognising foreign intruders or for limiting incestuous matings (O'Riain and Jarvis 1997, Burda 1995). One plausible hypothesis, regardless of actual olfactory acuity in social vs. solitary African mole-rats, is that the olfactory needs of these different social systems may have led to differential pace and intensity of evolutionary forces acting on olfactory genes.

Interestingly, a system of increased adaptive variability linked to bathyergid sociality is found at distinct MHC loci (Kundu and Faulkes 2004), which are thought to be 'assessed' via smell by conspecifics and thereby may play a role in directing mate choice decisions (Potts et al.1991, Penn and Potts 1998, 1999, Landry et al. 2001,

University of Cape Town

Olsson et al. 2003, Aeschlimann et al. 2003, Bonneaud et al. 2006). A comparison of four bathyergid species characterized by different social organisations – the solitary *Heliophobius argentocinereus*, the social *Cryptomys hottentotus hottentotus* and the eusocial *Fukomys damarensis and Heterocephalus glaber* – revealed high levels of adaptive variability at MHC class II DQα1 exon 2 alleles (Kundu and Faulkes 2004). Prevailing theory proposes that variation at class II MHC loci is the combined result of positive selection via heterozygote advantage, mate choice and parasite-host interactions (Doherty and Zinkernagel 1975, Jennions and Petrie 2000, Neff and Pitcher 2005, Reusch et al. 2001, Jacob et al. 2002, Milinski 2006). Accordingly, higher MHC diversity is correlated with increased individual and population fitness, as a consequence of pathogen pressure and/or selective mate choice. Because of the higher risk of parasite load in social groups, Kundu and Faulkes (2004) tested the prediction that the degree of sociality would positively correlate to the level of genetic polymorphism found; the resulting correlation was observed in comparisons of solitary and social mole-rat species.

Although there is no proven association between the OR genes identified in this thesis and the ability to detect MHC allele complements, it is nonetheless plausible to hypothesise that the bathyergid OR7 genes characterised here may play a role in kin recognition and mate choice; under such a scenario OR7 genes would be subject to similar evolutionary constraints as those of the MHC DQα loci, i.e. differential evolutionary rates between solitary and social ORs.

An argument for the involvement of olfaction in extra-colony member recognition and mate choice emerges from a study using non-coding genetic markers in the common mole-rat, *Cryptomys hottentotus hottentotus* (Bishop et al. 2004, 2007). *C. h. hottentotus* is a social, cooperative breeder that was traditionally believed to be strictly monogamous; the species shows highly developed individual recognition and incest avoidance behaviours, as well as marked aggressive behaviours to unrelated individuals (Bennett 1989, 1992, Spinks et al 2000). Based on both laboratory and mark-recapture studies, colonies of common mole-rats were found to be established by a single breeding pair of unrelated individuals (Bennett 1989, Bennett and Faulkes 2000). Nevertheless, a genetic survey by Bishop et al. (2004) using microsatellite markers revealed the occurrence of extra-colony paternity at unexpectedly high levels,

University of Cape Town

with 30% of offspring assigned to an extra-colony male. Furthermore, females were found to enhance offspring heterozygosity by mating with extra-colony males that were less related to them than their colony mates. However, male heterozygosity at the selected microsatellite loci did not influence their reproductive success. Bishop et al. (2007) proposed that the extra-pair mating behaviour of the common mole-rat may act to enhance offspring fitness by maximising the genetic compatibility of mates at selected genetic loci, such as the MHC. Given the substantial behavioural literature for mole-rats, the authors propose that the mechanism underlying the recognition of such genetic compatibility for African mole-rats is most likely to be olfaction (Bishop et al. 2007). Another study on parentage using microsatellite markers in the eusocial Damaraland mole-rat, *Fukomys damarensis*, provides further evidence for extra-colony paternity characterising the mating systems of social mole-rats (Burland et al. 2004). Thus, an olfactory mechanism similar to that proposed by Bishop et al. (2007) may be in place in other social African mole-rat species, where kin recognition and incest avoidance are fundamental to colony persistence.

University of Cape Town

4.1.2 Signatures of environmental niche specialisation on OR genomes

Increasingly, environmental niche specialisation is reflected in animal genomes, particularly among those genes involved in sensory physiology (e.g. Li et al. 2005, Seehausen et al. 2008, Zhao et al. 2009a). For example, a recent study on bat visual pigments by Zhao et al. (2009a) revealed a correlation between bat roosting ecology and the loss of function in short-wavelength sensitive (SWS) opsin genes, which are responsible for colour sensitivity to blue-violet light (Yokoyama 2000). In particular, Zhao et al. (2009a) found that tree-roosting bat lineages have retained functionality at SWS genes through strong purifying selection despite a long history of nocturnality (over 50 MY, Simmons et al. 2008), whilst in cave-roosters selection appears to have been relaxed, leading to a loss of functionality for SWS genes. Another example of the effects of environmental niche specialisation on genomes comes from the study of speciation through sensory drive (Schluter and Price 1993, Boughman 2002) in sympatric cichlid fish species from Lake Victoria (Seehausen et al. 2008). Lake Victoria is a highly heterogeneous habitat in terms of water clarity and ambient light (Levring and Fish 1956, Seehausen et al. 1997), and previous evidence for diversifying selection on the cichlid visual system had been linked to the adaptive

radiation of cichlids into several hundred species (Terai et al. 2006). In their study, Seehausen et al. (2008) found that long-wavelength sensitive (LWS) opsin genes, which are responsible for sensitivity to green-red light (Yokoyama 2000), have been shaped differentially among cichlid species; divergent selection according to the light regimes and water depths characteristic of the niches that species inhabit has led to a remarkable adaptive radiation in these fish. Another recent study on visual pigment evolution from Zhao et al. (2009b), compared rhodopsin genes (GPCRs) that are responsible for dim-light sensitivity (Yokoyama and Yokoyama 1996, Yokoyama and Shi 2000) across a number of mammalian species inhabiting both high- and low-light environments. Species coverage included several bats together with subterranean mole-rats, as well as a number of marine mammals that live in turbid water conditions. Although the rhodopsin gene was intact across all the species analysed, increased divergent selection was found across lineages that inhabit low-light habitats, suggesting an accelerated and independent evolution in these groups, in line with their ecological niche-specialisation (Zhao et al. 2009).

With regard to the evolution of OR subgenomes, a growing body of literature suggests that the environment has played a major role in shaping the vertebrate OR repertoire. The most comprehensive evidence for this assertion comes from a recent study by Hayden et al. (2010), consisting of a broad comparative survey of whole-OR-subgenome diversity across 50 mammalian species. In their study, Hayden et al. (2010) analysed the different proportions of functional OR genes and pseudogenes across all known OR gene families using a principal component analysis (Jolliffe 1986). The aim of Hayden et al.'s (2010) study was to identify which OR families have evolved differentially in mammals, and how this relates to the environmental niche that species inhabit. Interestingly, the study pinpoints a number of OR families that better explain the divergent OR diversity found between mammals from four main 'ecogroups': Terrestrial, Aquatic, Semi-aquatic and Volant (bats). To date, Hayden et al.'s (2010) study represents the most complete mammalian OR dataset in terms of species coverage and OR gene coverage; noteworthy, however, is that their study omitted the inclusion of subterranean mammals.

Other examples linking environmental niche specialisation to vertebrate OR evolution include the loss of OR functionality i.e. the high proportion of OR pseudogenes

University of Cape Town

reported in cetaceans, which appears to have coincided with the evolution of a marine lifestyle (Kishida et al. 2007). Interestingly, it appears that olfactory ability may vary among aquatic species, since both the numbers and proportions of functional OR genes are extremely variable across teleost fishes (Niimura and Nei 2005b, Niimura 2009). Within cetaceans, olfactory sensitivity appears to vary considerably between baleen whales (Mysticeti) and toothed whales (Odontoceti): the former have both higher proportions of functional OR genes, as well as a complex olfactory bulb which is absent in odontocetes, suggesting an increased olfactory ability in mysticetes (Thewissen et al. 2011). In snakes, viviparous species that have recently adapted to a marine lifestyle (8 MYA, Sanders et al. 2008) appear to have undergone extensive OR pseudogenisation, in comparison to both oviparous aquatic snake species, which rely on land for laying eggs, and terrestrial species (Kishida and Hikida 2010). This interesting trend observed across snake species supports the hypothesis that different OR genes are required in the aquatic and terrestrial environment (Niimura and Nei 2005, Kishida et al. 2007, Kishida and Hikida 2010). In line with a role for the environment in shaping OR subgenomes, Steiger et al. (2009) report expanded functional OR repertoires in two nocturnal bird species known to have a well-developed olfaction, when compared to their day-active closest living relatives.

University of Cape Town

4.1.3 Aims of this chapter

This chapter aims to test the roles of two non-mutually-exclusive hypotheses in the evolution of bathyergid OR7 genes, namely i) the Sociality Hypothesis and ii) the Ecogroup Hypothesis:

i) The Sociality Hypothesis: the social system of bathyergid mole-rats has influenced the evolution of OR genes resulting in differential evolution across bathyergid OR lineages, dependent on a species social system (solitary or social, with eusociality representing the extreme state of sociality).

ii) The Ecogroup Hypothesis: broad scale environmental-niche specialisation has significantly influenced vertebrate OR evolution, and this is predicted to be reflected in the proportion of functional OR genes characterising different mammalian 'ecogroups': Aquatic, Semi-aquatic, Terrestrial, Volant and Subterranean.

Alternatively, bathyergid OR genes have evolved under identical rates of molecular evolution across different taxonomic lineages or OR evolution proceeds independently of the environment in which species have evolved such that ratios of functional OR genes: pseudogenes are equivalent across the ecotypes analysed.

A number of different analytical approaches are used to test the respective roles of sociality and environment in shaping OR7 variation in the Bathyergidae. The relative contribution of mole-rat sociality is analysed using a phylogenetic framework, based on the bathyergid OR gene tree, where the two main characters of interest are mapped in turn onto the tree, corresponding to the social systems Solitary and Social. Because eusociality represents the extreme state of sociality, no division between 'social' and 'eusocial' will be discussed in this context, as emphasis is placed on the more general distinction between solitary and social bathyergid OR evolution. The role of the environment, on the other hand, is explored by integrating the bathyergid OR7 dataset developed in this study, into a broad analysis of orthologous mammalian OR7 genes (Hayden et al. 2010) that can now include the Subterranean ecogroup. In doing so, the

University of Cape Town

respective roles of two major aspects of bathyergid evolutionary ecology - sociality and environment - are explored.

4.2 Methods

4.2.1 Testing the role of sociality in shaping OR variation

The ratio of non-synonymous (dN) to synonymous (dS) substitutions at each codon site were calculated in order to explore the selective pressures acting on OR genes across the different bathyergid social groups. The ratios for two separate OR alignments were calculated using the SELECTON server (available at http://selecton.tau.ac.il/ - Doron-Faigenboim et al. 2005, Stern et al. 2007): one representing social species and the other representing solitary species.

To test the Sociality Hypothesis, three different types of analyses were performed using the Bathyergidae OR gene tree following (i) Ramm et al. 2008, (ii) O'Connor and Mundy 2009 and (iii) Mayrose and Otto 2011. The first step to these phylogenetic analyses is to distinguish between solitary and social Bathyergid species, and to label the terminal branches of the OR phylogeny according to the social status of the corresponding species. For all three analyses, the OR phylogeny reported in Chapter 2 was used (Figure 2.5 and Appendix III.2). The analysed data comprised all 119 unique bathyergid OR genes identified in this study, including both functional OR genes and pseudogenes. Stop codons from OR pseudogenes and alignment gaps were excluded from the OR alignment in Bioedit v7.0.8.0 (Hall 1999). OR genes from the genera *Bathyergus*, *Georychus* and *Heliophobius* were labelled as Solitary, while those belonging to *Cryptomys, Fukomys* and *Heterocephalus* individuals were labelled Social (Bennett and Faulkes 2000, Appendix I.9).

In the first analysis, levels of positive selection across specific bathyergid lineages were investigated following methodology first described by Ramm et al. (2008) in a study of genes encoding mammalian ejaculate proteins and involved in post-copulatory sexual selection. In their study, the authors sought evidence for positive selection on genes restricted to rodent lineages with large (or alternatively small)

University of Cape Town

relative testis sizes - their proxy for the intensity of sexual selection acting in each species. The authors used branch-site analyses (described in detail in Chapter 3, § 3.2.2 - Zhang et al. 2005) on phylogenetic gene trees that were 'partitioned' between species with relatively high levels of sperm competition (i.e. more intense post-copulatory sexual selection) and those with lower levels of sperm competition (i.e. less intense post-copulatory sexual selection). In Ramm et al.'s (2008) approach, phylogenies are 'partitioned' by mapping the phenotypes of interest onto the terminal branches of the gene trees, in their case into 'high' and 'low' sperm competition lineages; the method then tests the fit of models for positive selection onto these partitions. This procedure is considered to be very sensitive to positive selection, since in this case only two evolutionary rates - that of 'low' and 'high' sperm competition - are compared (Ramm et al. 2008). Nevertheless, it is important to bear in mind that this methodology is based on two assumptions: i) that the phylogenetic partitioning is biologically meaningful and ii) that the genes grouped together evolve under the same selective pressures (Zhang et al. 2005). In line with their expectations, Ramm et al. (2008) found that one of the seven rodent genes analysed carried a clear signal of adaptive evolution associated with sexual selection. In the same study, Ramm and colleagues extended their analysis - using the methods described above - to previously published data on primate semen coagulation proteins (Dorus et al. 2004), and concluded that positive selection is concentrated in primate lineages subject to high levels of sperm competition. Ramm et al.'s methodology was also recently used in a study of 13 mammalian ADAM sperm protein genes (i.e. containing A Disintegrin and A Metalloprotease domain, Wolfsberg et al. 1995) (Finn and Civetta 2010). In this study the authors also investigated a role for Darwinian evolution driven by mating systems and indeed found evidence of increased positive selection specifically in polyandrous primate lineages for one of the ADAM genes analysed.

In this thesis, Ramm et al's methodology is applied to the bathyergid OR gene tree partitioned between Solitary, and Social lineages as described above (Appendices I.9 and III.3). Analysis was performed using a branch-site test of positive selection (test 2 in Zhang *et al.* 2005) using CodeML (in PAML - Yang 1997, Yang 2007) as described in Chapter 3 (§ 3.2.2). Because there is no *a priori* hypothesis on which social system (i.e. solitary of social) would be subject to positive selection, two branch-site analyses were conducted. In the first branch-site test of positive selection, all the terminal

University of Cape Town

branches of the OR gene tree that belonged to social bathyergid ORs were labelled as 'foreground', to test whether social lineages carried a signal of increased selection when compared to the solitary ones. In the second analysis, the test was performed with the 'solitary leaves' of the tree labelled as foreground. With these branches defined as foreground, a LRT was performed on all pairs of nested models (null and alternative, parameter files reported in Appendix IV.1) and compared to a χ^2 distribution to determine significance. A Q-value was then calculated for each branch using the 'Q-value' software available at http://genomics.princeton.edu, in order to estimate the false discovery rate due to multiple testing (Storey 2002, 2003, Storey et al. 2004). When the LRT is significant (p-value < 0.05) and the false discovery rate is low (Q-value < 25%), the posterior probability of sites being under positive selection (dN/dS > 1) was then calculated using the Bayes Empirical Bayes (BEB) method (Nielsen and Yang 1998, Yang et al. 2005b).

One limitation of Ramm et al.'s approach (2008) is that phenotypic traits i.e. the 'social' or 'solitary' leaves of the phylogenetic tree, are assumed to have remained constant throughout the terminal branches. Phenotypes are applied to all portions of a branch thus ignoring the timing of the evolution of such phenotypes. In other words, phenotypic characters i.e. "instantaneous measures", are assumed to represent averages across entire branches and compared to dN/dS values which typically are true averages for an entire branch. It is not clear to what extent the comparison of such different data types may lead to erroneous inferences, but ideally equivalent data should be compared in a single statistical framework (Wong 2011).

To complement the branch site methodology described above, two additional tests, also based on maximum likelihood methods, were applied to the bathyergid OR phylogeny (O'Connor and Mundy 2009, Mayrose and Otto 2011), partitioned as described above. These two methods have the advantage of simultaneously modelling molecular and phenotypic evolution. For all character-states (i.e. phenotypes) at the terminal leaves of a phylogeny, probability distributions are estimated and assigned to internal nodes (Felsenstein 1981).

Both methods (O'Connor and Mundy 2009, Mayrose and Otto 2011) are based on model comparisons to detect significant associations between rates of sequence evolution and character-state evolution. In contrast to Ramm et al. (2008), these tests

University of Cape Town

do not distinguish between nonsynonymous and synonymous substitutions, but instead implement nucleotide-substitution models. Thus, significant associations between molecular evolutionary rates and phenotypic characters are not necessarily concomitant with changes in protein sequences. In these methods (O'Connor and Mundy 2009, Mayrose and Otto 2011), characters of interest are coded as binary values (0/1) - the 'social' and 'solitary' phenotypes in Bathyergidae. The models being compared in both maximum likelihood analyses are (1) a null model that assumes no correlation between evolutionary rates and character-states, and (2) an alternative model which allows phenotypic groups to have different evolutionary rates. The null and alternative models are compared via a likelihood ratio test (LRT), with rejection of the null model indicating a significant association between nucleotide substitution rates and phenotypic characters.

The approach of O'Connor and Mundy (2009) is based on a hybrid nucleotide substitution rate matrix, which governs the evolutionary process between combined genotypic-phenotypic states (i.e. for Bathyergidae 'nucleotide-sociality' states, e.g. A-Social, C-Solitary,…). Phenotype-specific site evolutionary rates are modelled via a set of scaling factors or weights (one per phenotype) that apply to a common background nucleotide substitution rate matrix. Only a fraction of sites are assumed to be subject to systematic variations in substitution rates, while the remaining sites are used to estimate the background substitution rates. In the null model, all scaling factors are forced to be equal to each other, which accounts for the presence of rate heterogeneity that might be independent of the phenotype. In the alternative model, they are directly interpreted as differential substitution rates that are specific to each phenotype. Hence, the associated LRT between the null and alternative model tests for differences in phenotype-specific substitution rates.

Mayrose and Otto (2011) proposed a Markov Chain Monte Carlo (MCMC) approach, which estimates phenotype-specific differences in substitution rates by integrating maximum likelihood estimations over phenotype evolutionary histories sampled along the gene tree. Phenotype-specific substitution rates are modelled by factors that systematically shrink or stretch portions of tree branches, according to their sampled phenotypic state. This implies that at a given time point (i.e. a tree node), all nucleotide sites are associated to a same phenotype. Significant correlation between

University of Cape Town

phenotype and substitution rates is computed using a bootstrap procedure, a permutation test in which the whole maximum likelihood estimation is performed multiple times on the original gene tree, but with sequence labels randomly shuffled.

A fundamental difference between the O'Connor and Mundy (2009) and Mayrose and Otto (2011) methods lies in the way changes in phenotype affect substitution rates along a sequence. Under O'Connor and Mundy's model, the phenotype-genotype evolutionary process of each nucleotide site is independent. On the other hand, under Mayrose and Otto's model, all sites are assumed to evolve in a phenotype-dependent manner, described by each sampled phenotype tree. Thus, the O'Connor and Mundy model (2009) is more suitable to situations when the rates of molecular evolution of some, but not all, sites are associated to phenotypic character-states. This is similar to the way in which codon-based models detect positively-selected sites (e.g. the branch-site test of positive selection as implemented in Ramm et al. 2008). Mayrose and Otto's method (2011), in contrast, aims to detect lineage-specific mutational effects that affect all sites.

The maximum likelihood methods described above were applied to a number of different datasets. In their study, O'Connor and Mundy confirmed the results from Dorus et al. (2004, also in line with Ramm et al.'s 2008 findings described above) on primate semen coagulation proteins, *SEMG1* and *SEMG2*. A strong association is supported between the rates of molecular evolution and mating system for *SEMG2*, consistent with a correlation between mating system and the intensity of sexual selection on this gene. O'Connor and Mundy also applied their model to a previously described dataset from the sperm ligand zonhadesin gene (ZAN), and for which a negative correlation was found between dN/dS and sexual size dimorphism in primates (Herlyn and Zischler 2007). O'Connor and Mundy's method however failed to confirm this correlation, perhaps due to the differences in which sexual selection is measured in the two different studies (binary classification versus a continuous measure of sexual body mass dimorphism), a lack of power associated with the method, or the inaccuracy of the original study by Herlyn and Zischler (Wong 2011). Mayrose and Otto, on the other hand, applied their method to a dataset of mitochondrial genes from saline and freshwater species of crustaceans (genus *Daphnia*, Colbourne et al. 2006). In line with their expectations, they successfully

University of Cape Town

detected an effect of habitat salinity on nucleotide substitution rates, with higher substitution rates in halophilic daphniids (inhabiting saline waters) being ascribed to the mutagenic effect of high salt concentrations and/or levels of UV radiations in saline habitats (Hebert et al. 2002). Interestingly, Mayrose and Otto applied O'Connor and Mundy's method to the same dataset (Colbourne et al. 2006), but no correlation between habitat and molecular evolutionary rates was found (Mayrose and Otto 2011). This may suggest that saline habitats cause a genome-wide increase in the nucleotide substitution rates of daphniids, rather than an association which may be limited to a fraction of sites only, and which would have been detected by O'Connor and Mundy's method (Wong 2011).

Here, the O'Connor and Mundy (2009) and Mayrose and Otto (2011) methods were applied to the bathyergid OR phylogeny partitioned into Solitary and Social terminal branches. O'Connor and Mundy's analysis was performed using the HyPhy scripts made available by the authors in supplementary data at http://bioinformatics.oxfordjournals.org/content/25/12/i94/suppl/DC1, which were adapted to the bathyergid OR dataset (Appendix IV.2). Mayrose and Otto's analysis was performed with the program traitRate (available at http://www.zoology.ubc.ca/~mayrose/cp/traitRate/), following the authors' parameter choices (Appendix IV.3). Due to the difference in taxonomic coverage in the phenotypic groups (88 social vs 31 solitary sequences), the complete procedure was run twice, using either group as the reference (I Mayrose pers. comm.). The intial run used 200 stochastic mappings; significance was computed with 500 bootstrap runs, with 200 stochastic mappings each.

University of Cape Town

4.2.2 The subterranean 'Ecogroup' as a driver of bathyergid OR diversification

The role of the subterranean environment as a dominant driver of OR evolution across the Bathyergidae (Ecogroup Hypothesis) was tested by comparing the ratios of functional OR genes:pseudogenes across ecotypes (following Hayden et al. 2010). Proportions of (non)functional OR genes are traditionally used as markers of olfactory acuity, with increased OR pseudogenisation being an indicator of the olfactory decline of a species (Godfrey et al. 2004, Ache & Young 2005, Kishida 2008, Keller and Vosshall 2008). As detailed in Chapter 3, the proportions of (pseudo)genes also

reflect aspects of the evolutionary history of ORs (Gilad et al. 2004, Kishida et al. 2007) since, in the context of 'birth and death' evolution, duplicated genes that escape adaptive selection or fail to acquire new functions are predicted to undergo pseudogenisation (Kambere and Lane 2007). Therefore, OR pseudogenes either accumulate in the genome, become unidentifiable, or are deleted from the genome (Kambere and Lane 2007).

Using this framework, Hayden et al. (2010) generated a dataset with OR functional genes:pseudogenes ratios from whole-OR-subgenome data for 50 mammalian species, representing the most complete mammalian OR dataset to date in terms of both species and OR gene coverage. The species analysed covered a range of environmental niches, namely Terrestrial, Aquatic, Semi-aquatic and Volant (i.e. bats). A Bayesian phylogenetic analysis was performed to classify OR genes into gene families, and the 17 'traditional' OR families were recovered (Chapter 2, Glusman et al. 2000a); the following families were found to group together OR 2/13, OR 1/3/7, OR 5/8/9. Unfortunately, the raw sequence data used by Hayden et al. (2010) for the Bayesian analysis have not been made publicly available; only the numbers and ratios of functional OR genes and pseudogenes are available in their published study. Hayden et al. (2010) used a principal component analysis, based on the different proportions of functional ORs and pseudogenes across gene families, to compare data from the different 'ecogroups', and in so doing identify the OR families that explain most of the variation between these groups.

Sequence similarity and phylogenetic data (Chapter 2) revealed that all the mole-rat OR genes identified in this study fall into one main OR family, namely OR7. Following Hayden et al. (2010), this particular gene family belongs to a broader mammalian OR grouping of three OR families, that of the OR 1/3/7. In this chapter, published data on the numbers of functional OR 1/3/7 genes and pseudogenes, as well as their relative proportions as reported in Hayden et al. (2010), are used together with data from the 14 Bathyergidae species analysed in this study to test the Ecogroup Hypothesis. The dataset used is reported in Appendix I.8, and consists of taxonomic information, as well as numbers and proportions of functional OR genes and pseudogenes per species within OR 1/3/7. Furthermore, species are classified

University of Cape Town

according to five 'ecogroups': Aquatic, Semi-aquatic, Subterranean, Terrestrial and Volant (Appendix I.8).

OR1/3/7 pseudogene ratios were plotted across all species within each ecogroup and the mean percentage of pseudogenes and associated standard error were calculated in R (R Development Core Team, 2008, http://www.R-project.org). To test for pairwise differences in the distributions of pseudogene proportions between each ecogroup, a non-parametric Wilcoxon-test (Wilcoxon 1945) was applied, and the Benjamini & Hochberg, BH, correction for multiple testing was applied (Benjamini and Hochberg 1995).

University of Cape Town

4.3 Results

4.3.1 Sociality results

In this chapter, evidence for whether episodic positive selection has acted differentially on OR7 genes across specific bathyergid lineages is explored. An overall view of dN/dS ratios across bathyergid ORs, partitioned in Solitary and Social species, is reported in Figure 4.1. Although the overall trend of dN/dS ratios is similar across the two social groups, certain amino-acid locations display marked differences in their ratios, with either the solitary or social bathyergid ORs having higher dN/dS values.

To test the Sociality Hypothesis, the bathyergid OR gene phylogeny was explored using three different tree-based methods (Ramm et al. 2008, O'Connor and Mundy 2009, Mayrose and Otto 2011). All three analyses are based on likelihood ratio tests, for which we report the estimated log-likelihood differences and respective p-values. None of the methods identify a significant correlation between social phenotypes and either positive selection (LnL difference=0, p=1; following Ramm et al. 2008), or genotypic evolutionary rates (LnL difference=0, p=1; O'Connor and Mundy 2009, LnL difference=8.4, p=0.24; Mayrose and Otto 2011). For example, Figure 4.2 shows an histogram of the bootstrap distribution of 200 log-likelihood differences, as computed for testing significance in the analysis following Mayrose and Otto 2011. The log-likelihood difference and p-value obtained for the original Bathyergid OR7 data are indicated in blue, while the dark gray line indicates the threshold value required to infer a correlation between genotypic evolutionary rates and social phenotypes with a 0.05 p-value.

University of Cape Town

Figure 4.1 dN/dS ratios at OR codon sites in solitary vs social Bathyergidae dN/dS ratios calculated across codons from OR alignments of solitary (orange) and social (blue) bathyergids. To visualise the dataset that is used to test for the Sociality Hypothesis, OR pseudogenes are included in this analysis, but stop codons and codons containing alignment gaps are discarded; thus, amino-acid positions on the X axis do not correspond to the positions indicated in Figure 2.3.

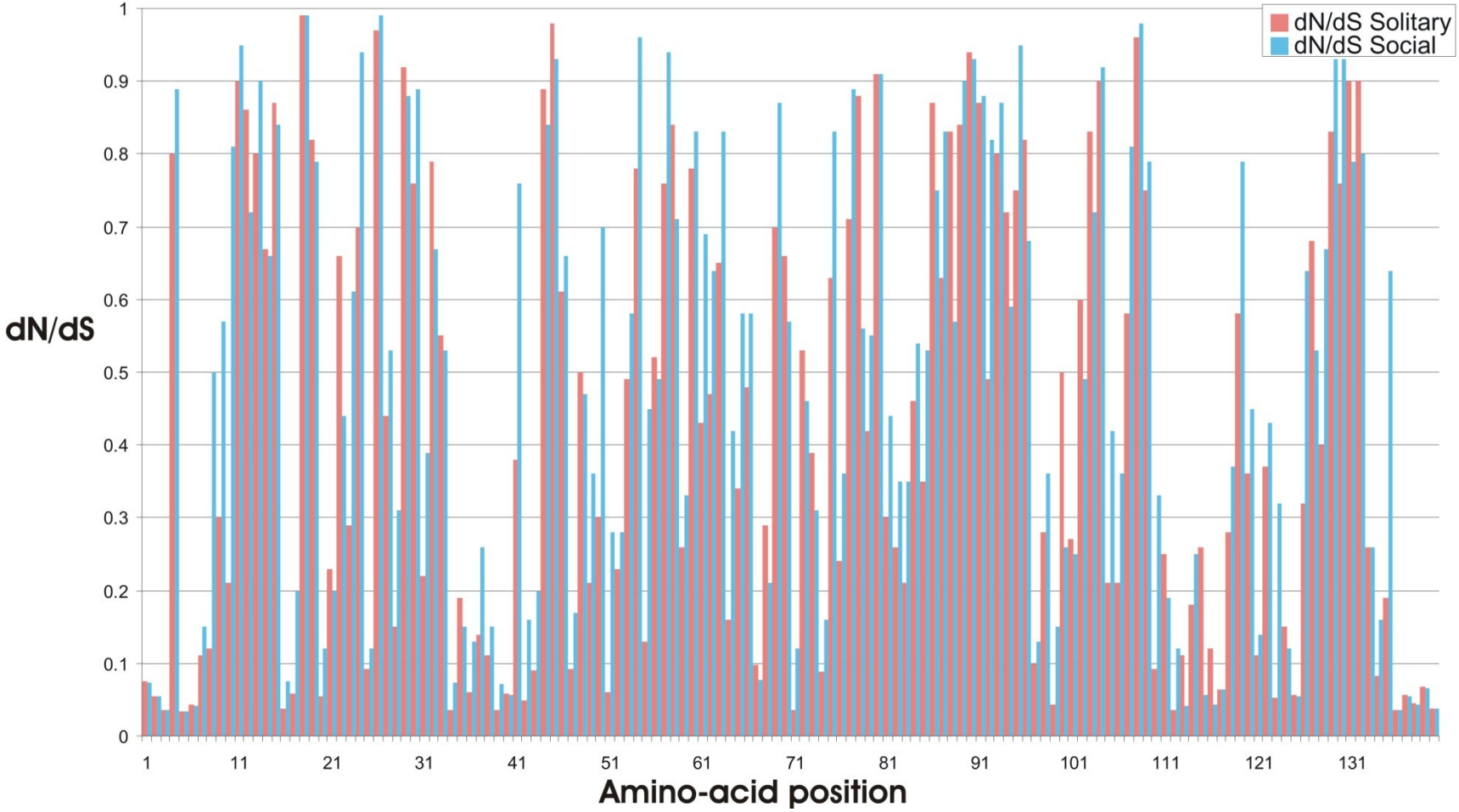
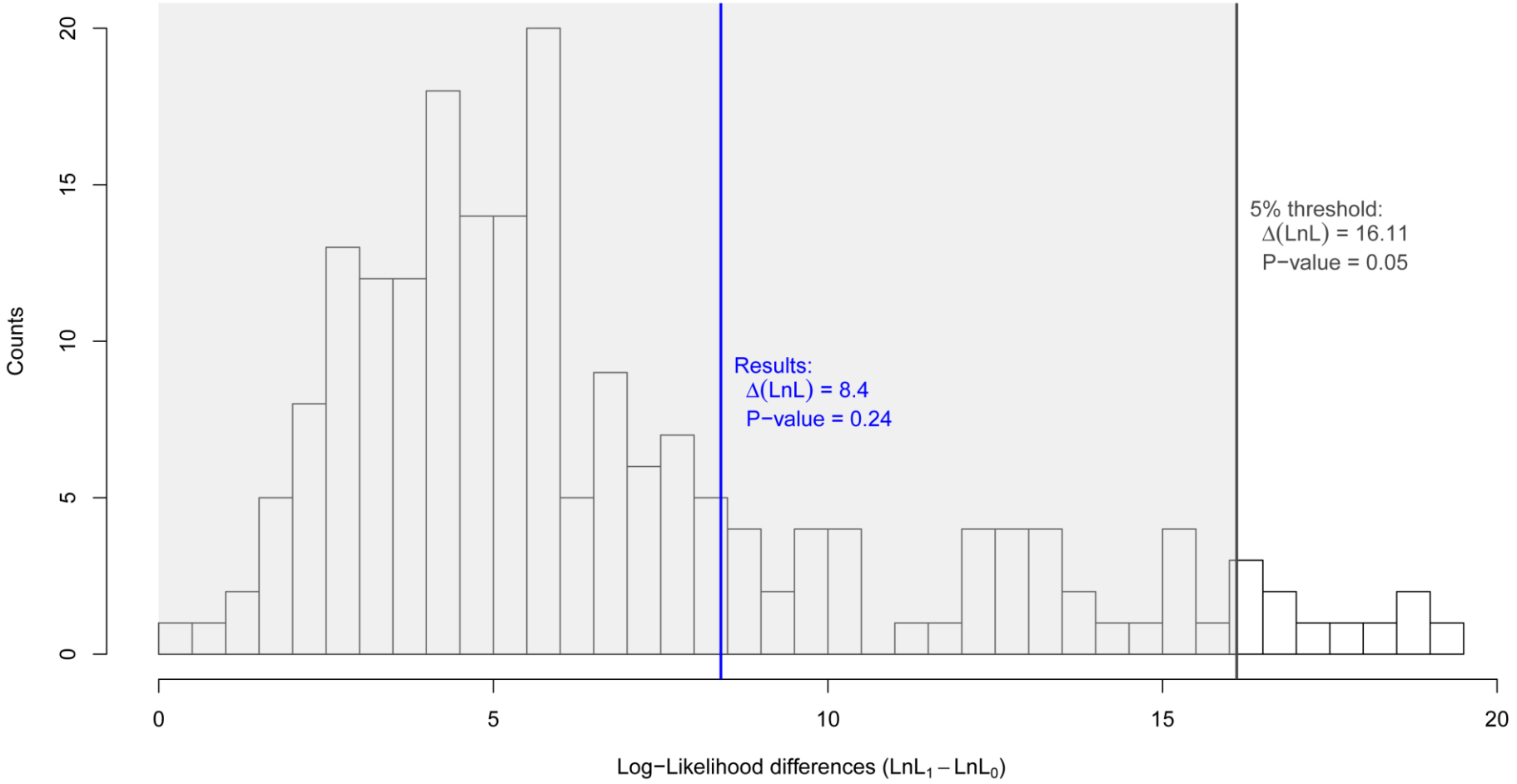


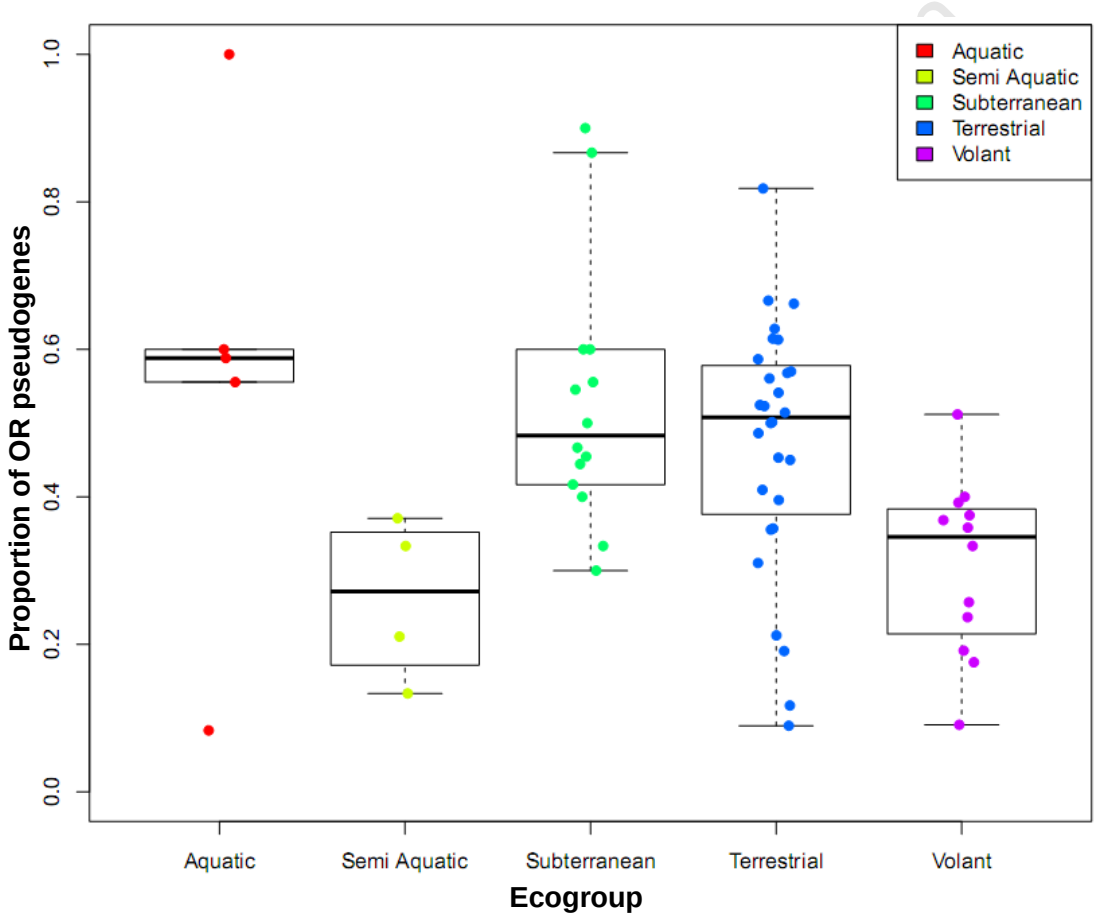
Figure 4.2 Bootstrap distribution of log-likelihood differences between the alternative and null models from Mayrose and Otto 2011 Observed ratio and p-value are indicated in blue; the 0.05 significance threshold is in dark gray; the light gray area shows the range of values that do not reach this threshold.



4.3.2 Ecogroup results

In order to establish the role of the environment in shaping OR diversity, OR 1/3/7 genes were compared across a suite of mammalian species occupying the full spectrum of ecological habitats. Following Hayden et al. (2010), the different proportions of OR 1/3/7 pseudogenes were considered within each Ecogroup, introducing the 'Subterranean' group to the analysis. Proportions of OR pseudogenes within Ecogroups are reported in Figure 4.3.

Figure 4.3 Proportions of OR 1/3/7 pseudogenes across Ecogroups The mean percentage of pseudogenes and standard error are indicated for each Ecogroup.



A Wilcoxon rank-sum test (Wilcoxon 1945) was used to identify which ecogroups differ significantly with respect to their proportions of (non)functional OR1/3/7. Results are detailed in Table 4.1. Significant differences were found between Subterranean and Semi-aquatic, Subterranean and Volant, and Terrestrial and Volant Ecogroups only; a nearly significant value differentiates the Terrestrial from the

Semi-Aquatic Ecogroup (p = 0.06, Table 4.1). Thus, whilst the subterranean environment potentially contributes to the evolution of the observed differences in OR 1/3/7 ratios, this data set fails to provide unequivocal support for the Ecogroup Hypothesis; the limitations associated with this approach are discussed in the next section.

Table 4.1 Pairwise comparisons between Ecogroups using Wilcoxon rank sum test (p-value adjustment method: BH) Ecogroups that differ significantly ($p < 0.05$) are highlighted in yellow.

	Aquatic	Semi-Aquatic	Subterranean	Terrestrial
Semi-Aquatic	0.317	-	-	-
Subterranean	0.505	0.036	-	-
Terrestrial	0.483	0.06	0.81	-
Volant	0.127	0.518	0.011	0.014

Within the Subterranean ecogroup, the proportions of OR pseudogenes are variable, as summarised in Table 4.2. The comparatively high proportion of OR pseudogenes in the naked mole-rat, *H. glaber*, will be taken into consideration when discussing the results of this section.

Table 4.2 Proportions (%) of OR pseudogenes across bathyergid species, genera and social groups The % of pseudo-ORs in bathyergid genera and social groups are calculated as the mean % of pseudo-ORs across species within each genus or social grouping, respectively.

% OR7 pseudogenes					
Species		***Genus***		***Social structure***	
Bathyergus janetta	60.0				
Bathyergus suillus	55.6	*Bathyergus*	57.8		
Georychus capensis	44.4	*Georychus*	44.4		
Heliophobius argentocinereus	30.0	*Heliophobius*	30.0	**Solitary**	**45.5**
Cryptomys hottentotus hottentotus	40.0				
Cryptomys hottentotus natalensis	50.0				
Cryptomys hottentotus pretoriae	60.0	*Cryptomys*			
Fukomys amatus	45.5				
Fukomys anselli	46.7				
Fukomys bocagei	54.5				
Fukomys darling	90.0				
Fukomys mechowi	33.3			**Social**	**53.2**
Fukomys damarensis	41.7	*Fukomys*	51.9		
Heterocephalus glaber	86.7	*Heterocephalus*	86.7	**Eusocial**	**66.7**

University of Cape Town

4.4 Discussion

The remarkable features of the Bathyergidae include i) that mole-rat societies embrace the complete array of social systems described in mammals, and ii) that their highly adapted sensory biology allows full and highly successful exploitation of the rather extreme subterranean niche. Both these features require enhanced olfaction, which is known to be functionally underpinned by a broad and variable OR gene repertoire (Niimura and Nei 2006, Kishida 2008). Indeed, this study presents clear evidence of positive selection acting on mole-rat OR7 genes (Chapter 3), suggesting that adaptive evolution has operated to increase olfactory acuity during the evolution of extant bathyergids. Here, the unique features listed in points i) and ii) are investigated for possible roles in a scenario of adaptive evolution.

4.4.1 Sociality as a driver of natural selection on olfactory receptor genes

The contribution of a mole-rat species' sociality (or solitariness) in shaping OR7 diversity was assessed using three tree-based methods following i) Ramm et al. (2008), ii) O'Connor and Mundy (2009) and iii) Mayrose and Otto (2011). All three analyses failed to detect a significant correlation between the social phenotype of bathyergid species and positive selection (i) or the rate of molecular evolution (ii & iii), respectively. Therefore, the bathyergid OR7 dataset does not provide support for the Sociality Hypothesis. Despite this lack of evidence, the Sociality Hypothesis cannot be discarded unequivocally because a number of factors, principally linked to the type of data considered, can account for such a result.

One of the most important factors limiting the Sociality Hypothesis test is the relative size of the OR7 dataset considered. As described in Chapter 2, the number of OR7 genes isolated in this study, although likely to be representative of overall OR7 variation, is certainly limited, and the extent to which the OR7 subfamily has been characterised in Bathyergidae is uncertain. It could be argued that a cross-species and cross-genes investigation of genetic variability in such a large gene family inevitably leads to incomplete datasets, particularly so when data collection relies on PCR-sequencing techniques as opposed to *in silico* data mining searches from whole-

University of Cape Town

genome estimates. However, *in silico* data mining searches are, of course, limited to single or indeed model species. Notwithstanding, the results presented here do not invalidate an investigation of the hypotheses tested in this chapter, despite a limited and patchy dataset; this is particularly so given that mole-rats are truly novel non-model species (but see recent literature arguing for the naked mole-rat as a model in human aging research, Buffenstein 2005, Holmes and Kristan 2008).

In a recent review on the evolution of reproductive tract proteins (Wong 2011), the authors stress that it is indeed generally reasonable to expect natural selection to act on different gene targets in different species. This is supported by findings of accelerated evolution and turnover of selected genes in some species (e.g. reproductive tract genes in *Drosophila*, Holloway and Begun 2004, Haerty et al. 2007), and lineage-specific observations of positive selection on specific proteins (e.g. Turner et al. 2008, Finn and Civetta 2010). Thus, although a signal of differential evolution linked to character-state may not be evident in many single-gene single-species studies, it could be revealed by cross-gene and cross-species studies. For example, in the *Drosophila* fruitfly genus, the intensity of post-copulatory sexual selexion (PCSS) is thought to be higher in species from the *D. repleta* group when compared to the *D. melanogaster* group because of the higher re-mating rates in the former group (Markow 1996, Markow and O'Grady 2005). The increased PCSS in *D. repleta* has been confirmed by a number of studies on reproductive-tract genes, which reveal higher duplication and dN/dS rates in *D. repleta* species when compared to species from the *D. melanogaster* group.

Another interesting point raised by Wong (2011) with regards to identifying possible targets of positive selection among reproductive genes, is that gene characterisation should preferably extend to several individuals per species analysed. This, in order to be able to detect signals of selection which may have acted on one species only, which otherwise may not be consistent enough to be detectable across a phylogeny (Wong 2011).

Both points raised by Wong (2011) justify the *a priori* investigation of selective forces that more than likely have acted across species and OR genes in the Bathyergidae. Ideally, sequences from more OR7 genes across all the bathyergid species analysed, and across several individuals per species should be investigated and indeed can be carried out using standard PCR-sequencing techniques (as used here); however, this approach is both time- and budget-consuming and was therefore

University of Cape Town

impossible to implement in the context of my PhD. Alternatively, ORs could be investigated via data mining searches, which are however restricted to those species whose genomes have been sequenced. In the last weeks of completion of this thesis, a preliminary 20x coverage whole-genome sequence of the naked mole-rat was published online (*Naked Mole-Rat Genome Resource 2011*, http://naked-mole-rat.org); whilst un-annotated, it does nonetheless represent a comprehensive resource with a built-in BLAST search tool. The availability of the whole-genome sequence data for this species is extremely exciting and will allow for specific OR data mining in the near future. Specifically, in the context of the Sociality Hypothesis, data mining could reveal what proportion of the naked mole-rat OR7 repertoire was amplified in this study. This would give a realistic estimate of how much more OR7 sequencing would be required to generate a comparable data set across Bathyergidae species, in order to accurately test for the Sociality hypothesis.

It is also possible that the OR7 family characterised in this thesis is not responsible for the detection of odours that are relevant in a social context, but rather for recognising allelochiemic substances. If this is the case, the bathyergid OR7 repertoire would not be useful to testing the Sociality Hypothesis, as an OR family responsible for detecting semiochemicals would preferably need to be targeted for this purpose. The search for such a gene family is however limited by the current difficulty of functionally characterising OR receptor-ligand interactions *in situ* (Saito et al. 2009).

Further caveats in testing the Sociality Hypothesis come from the analytical methods used. All three methodologies applied to the bathyergid OR dataset to address the Sociality Hypothesis were originally designed for the analysis of single-gene datasets. Here, there are multiple OR genes, amd so evolution for multiple genes is being mapped over the species tree which incorporates the transitions between solitary and social. TFor the Ramm et al. (2008) method, this leads to extra uncertainty in the phenotypic evolution along the tips of the tree that would not be present with a single locus, because one cannot be sure that all of the orthologous ORs are present in the dataset. So, for a tip leading to an OR of a certain species, the true orthologue in other species may not be in the dataset and, in effect, the length of the tip is being oversetimated. This has the overall effect of increasing noise in the analysis. Similarly, for the two other methods which use explicit models of genotype-

University of Cape Town

phenotype evolution (O'Connor and Mundy 2009, Mayrose and Otto 2011), the data structure violates the assumptions . The effects of such violation are unclear, but it would likely decrease the power of the analysis.

4.4.2 Environmental niche specialisation and OR make-up across 'ecogroups'

Following Hayden et al. (2010), the Bathyergidae OR7 dataset characterised in this study was integrated into a broad analysis of orthologous mammalian OR genes to test for the role of environmental niche-specialisation in mole-rat OR diversification (the Ecogroup Hypothesis). Proportions of (non)functional ORs across OR1/3/7 families reveal that the Subterranean ecogroup differs significantly from the Volant and Semi-aquatic groups, but is not significantly different from the Terrestrial and Aquatic groups (Figure 4.3, Table 4.1). Thus, data from the mole-rat OR7 family does not provide unequivocal support for the subterranean environment *per se* as a strong driver of evolutionary pattern across OR7 genes within the framework of the Ecogroup Hypothesis.

Again, it is important to indicate the intrinsic limitations of the methodology used in this analysis. The approach used is dependent on the type of dataset used, the different evolutionary mechanisms that shape and determine the ratios of OR genes: pseudogenes in the gene family under study, and the different ages of the mammalian families that are compared.

A number of observations can be made on the nature of the dataset used to test for the Ecogroup Hypothesis. As for the Sociality Hypothesis, it is important to bear in mind that, although the OR repertoire characterised in this thesis is likely representative of overall OR7 diversity in Bathyergidae (Chapter 2), it is nonetheless very limited in comparison to the whole-genome estimates used in the Ecogroup analysis (Hayden et al. 2010).

With respect to the lack of significant difference between the Aquatic and Subterranean data set, it should be noted that the Aquatic ecogroup is based on a small sample of species (i.e. only five species) which includes two species with large differences in their proportion of OR 1/3/7 pseudogenes: 8% for the sea otter vs 100%

University of Cape Town

for the pilot whale, where only one pseudogene is in fact described in the species and no functional genes have been identified to date (Figure 4.3, Appendix I.8). These two contrasting OR 1/3/7 compositions distort the average mean proportion of pseudogenes in the Aquatic ecogroup resulting in a large standard error. Thus, it is possible that the undetected difference between the Aquatic and other ecogroups is due to the extreme range of pseudogene fractions in the Aquatic 'gene pool' analysed. Undeniably, a complete assessment of OR 1/3/7 genes across all species analysed would be needed for unambiguous results.

With respect to the use of (non)functional OR gene ratios to differentiate between ecogroups, it has been argued that such proportions are not always an accurate estimator of olfactory acuity and therefore may not necessarily be good indicators in comparative studies across species (Nei et al. 2008). In this scenario it is the extent of the functional OR subgenome of a species that determines its olfactory sensitivity, and not the proportion of functional OR genes, because ultimately a greater number of ORs will determine the ability to detect and discriminate more odorants. For example, a recent comparative study of bird species reveals that, despite having similar proportions of OR (pseudo)genes, the nocturnal kakapo (*Strigops habroptilus*) and brown kiwi (*Apteryx australis*) have greater numbers of functional ORs in their repertoires than their closest diurnal relatives (Steiger et al. 2009). This is a typical example for a role of environmental niche specialisation in shaping OR genomes, where the increased olfactory sensitivity of the nocturnal birds analysed (Steiger et al. 2009), is not reflected by a proportionally increased fraction of functional OR genes. Another interesting example, within OR 1/3/7 genes, is seen in the house mouse and the lesser horseshoe bat who have identical proportions of OR pseudogenes (9% in both species, Appendix I.8): the former has 132 functional genes while the latter only has 10 (Hayden et al. 2010), suggesting that olfactory sensitivity, at least that conveyed by OR 1/3/7 receptors, is likely higher in the mouse.

A further limitation of the analytical approach used here comes from the heterogeneous taxonomic coverage and the different ages of the taxa being compared across ecogroups. The species coverage in the Terrestrial ecogroup, for example, spans across four superorders of mammals, with 28 species from more than 20 different families, and extremely variable lineage ages e.g. Muridae 31 MY (Adkins

University of Cape Town

et al. 2001), Elephantidae 25 MY (Rohland et al. 2007), Hominidae 21 MY (Chatterjee et al. 2009), Canidae 12 MY (Bardeleben et al. 2005). On the other hand, only a single relatively ancient mammalian family - Bathyergidae 49 MY (Nedbal et al. 1994, Blanga-Kanfi et al. 2009) - represents the subterranean ecogroup. Ideally, more balanced species coverage across ecogroups, considering only those taxa with similar age would be needed to test more accurately for the Ecogroup analysis. Intuitively, older species may have accumulated a greater proportion of pseudogenes simply as a function of time, because continuous 'birth and death' evolution would theoretically lead to an increase of OR pseudogenes which are essentially neutral (Li et al. 1981, Gilad et al. 2003). Even though OR pseudogenes should eventually become unidentifiable due to accumulated mutations, it is important to remember that ORs classified as 'non-functional' may play a regulatory role in gene expression, as suggested by Zhang et al.'s (2007b) finding that 67% of human pseudogenes are in fact transcribed. This may explain the persistence of OR 'pseudogenes' in the genome over long periods of time.

In this context, it is interesting to note that the naked mole-rat, *H. glaber*, the oldest species in Bathyergidae (Chapter 1, Figure 1.7), has an extremely high proportion of pseudogenes within OR7 (86.7%, Table 4.2). Again, the recent availability of the whole-genome sequence data for the naked mole-rat (*Naked Mole-Rat Genome Resource 2011*, http://naked-mole-rat.org) will allow for specific OR data mining in the near future in the context of the Ecogroup Hypothesis. This can be used to verify whether the high proportion of OR pseudogenes identified in this thesis is a real trend in *H. glaber* OR7 genes, perhaps linked to the old age of the lineage, or simply a function of limited sampling. Furthermore, the characterisation of the entire OR subgenome in the naked mole-rat will reveal which OR families have diversified the most. As with the Sociality Hypothesis, the appropriateness of the present dataset to test the Ecogroup Hypothesis ultimately depends on the type of odorants that are detected by bathyergid OR7 genes. If these odorants are not predominantly typical of the underground environment, the OR7 gene repertoire may not be suitable to testing the Ecogroup Hypothesis because the genes would not be subject to differential habitat-driven evolution. Expanded OR families represent the ideal targets for use in the Ecogroup analysis because under this hypothesis ORs responsible for the detection of environmental odorants are predicted to have experienced a wave of selective pressure during the transition to the subterranean environment. Complete

University of Cape Town

sequence information will also allow for the design of bathyergid specific oligonucleotide primers to extend the search for target OR families to other African mole-rat species.

To conclude, neither hypothesis tested in this chapter is firmly supported, possibly due to the limitations of the experimental procedure used. A role for sociality and the environment in shaping OR variation cannot however be excluded. Olfactory requirements without doubt differ between solitary and social bathyergid species, due to the fundamental differences in lifestyles which, for example, require social species to optimise kin recognition (e.g. to avoid incestuous matings, Burda 1995). The observed tendency of the subterranean environment to shape OR diversity is only partly consistent with Hayden et al.'s (2010) conclusions that natural selection, via niche-specific adaptation, shapes OR subgenomes. Nevertheless, it is reasonable to propose that the olfactory requirements of species that inhabit such diverse ecogroups are indeed different and may be reflected in other OR gene families. The necessity to detect either airborne or water-soluble odorants is the most tangible reason why the OR repertoires of terrestrial and aquatic species should differ (Freitag et al. 1998, Niimura and Nei 2006, Nei et al. 2008). The subterranean environment, on the other hand, presents different challenges including the absence of visual cues and limited auditory cues, requiring fossorial species to compensate with enhanced olfaction, reflected in a distinctive and large OR subgenome.

University of Cape Town

Chapter 5: Synthesis and Conclusion

Gene duplication and loss is undoubtedly a potent source of evolutionary innovation and genomic analysis provides powerful tools for the study of evolution. The genomic analysis approach can provide insight into the evolution of a number of intrinsic and behavioural traits of wild species, that complement lengthy and challenging field observations. Accordingly, the techniques of molecular biology are particularly suited to the study of species that are difficult to access in their natural habitat, such as the African mole-rats. This thesis investigated the molecular evolution of olfactory genes in the subterranean Bathyergidae, which are known to rely on olfactory cues for a number of vital tasks (Faulkes 1990, Judd and Sherman 1996, Heth et al. 2002a, 2002b, 2004, Lange et al. 2005). Analyses based on the genetic make-up of a representative portion of the OR subgenome reveal that the evolutionary history of Bathyergidae and the evolution of olfaction are intermingled.

Within the subterranean environment, habitat specialisation to a range of humidity and soil-hardness clines has played an established role in the evolution of bathyergid sociality, as detailed in the well-supported Aridity-Food Distribution Hypothesis, AFDH (Chapter 4, Jarvis and Bennett 1991, Jarvis et al. 1994, Faulkes 1998). According to the data in this thesis, the 'subterranean' environment fails to emerge as an unequivocal driver for the observed diversification of the OR7 subgenome family (Ecogroup Hypothesis, Chapter 4; see Figure 4.3 and Table 4.1). The role of the environment in shaping the olfactory abilities of species is implicitly invoked in numerous studies on the evolution of the OR multigene family (Glusman et al. 2001, Niimura and Nei 2005b, Niimura 2009). The rationale behind this hypothesis is that the olfactory needs of organisms which inhabit varied and specialised ecological niches will vary according to the habitats in which species evolve, and this will be reflected in their OR repertoires. A striking example of this is the extensive diversification that the vertebrate OR family underwent after the transition from the Aquatic to the Terrestrial environment at the end of the Devonian period some 370 MYA (Niimura and Nei 2005b); a 10-fold increase in the number of OR genes in

University of Cape Town

terrestrial vertebrates compared to teleost fish in all probability reflects the necessity to detect a novel and broad range of airborne odorants (Glusman et al. 2001, Niimura and Nei 2005b). This sensitivity of the OR subgenome to respond to an ever-changing environment is also observed at smaller time-scales: the OR repertoires of humans, for example, reflect the recent history of human migration and the variability in environments and diets to which different human populations adapted (Gilad and Lancet 2003, Menashe et al. 2003). In this context, the fact that this study fails to determine a definite role for the specialised subterranean niche in shaping mammalian OR7 repertoires, is likely a consequence of the limitations of the experimental procedure used and the OR7 subgenome itself (Chapter 4). Nevertheless, such a role for the environment is not excluded.

Similarly, the data in this thesis is not consistent with the shaping of the bathyergid OR7 subgenome by the varied social systems of African mole-rats. There is no clear evidence that OR7 genes have been subject to different rates of molecular evolution depending on the social phenotype of their associated species ('solitary' or 'social' - Sociality Hypothesis, Chapter 4). In light of recent studies on bathyergid sensory physiology (Cris et al. 2003, Park et al. 2003, Rosen et al. 2007, Holmes et al. 2008), the results from this thesis nevertheless represent a starting point to explore the likely contribution of sociality in driving sensory specialisation, by setting a methodological framework in which the question can be addressed on a more complete dataset. The majority of neurobiological studies on bathyergids have focused on the naked mole-rat, *Heterocephalus glaber*, which has emerged as a very interesting model organism. Among Bathyergidae, *H. glaber* represents the oldest lineage and presents a number of unique features that are not all shared by other bathyergids, thus lessening the generalisation of findings from this single species to the rest of the extant Bathyergidae. Here, the comparison of ORs across bathyergid species emerges as a novel informative tool for the study of socially-related traits in the family. Thus, despite the rejection of the Sociality Hypothesis (Chapter 4), one of the major strengths of this research is the broad array of species that are compared with an accessible and easily reproducible methodology.

Olfaction clearly influences sociality and environmental niche-specialisation by conferring a fundamental flexibility in olfactory responses, which potentially allow

University of Cape Town

for the exploitation of an array of habitats and social systems. The different mechanisms of natural selection identified across mole-rat ORs (Chapter 3) represent the most striking evidence of olfactory 'plasticity' (i.e. 'flexibility', 'sensitivity' or 'responsiveness') and are consistent with the biological framework in which OR genes function. To further explore the evolution of ORs in Bathyergidae, one would ideally integrate the present dataset with a range of orthologous OR7 genes from closely-related hystrycognath rodents, encompassing different ecological and social niches. Species coverage would typically mimic the recent studies of South American Hystricognathi (Caviomorpha), which have deciphered caviomorph phylogenies by depicting interesting patterns of molecular evolution across several non-OR genes (Wlasiuk et al. 2003, Castillo et al. 2005, Opazo 2005, Opazo et al. 2008, Lessa et al. 2008, Parada et al. 2011), possibly in relation to specific caviomorph life-history traits (Opazo et al. 2008). The addition of ORs from other hystricognath species does however depend on the availability of samples, which usually requires prolonged and challenging field collection.

Smell is intrinsically linked to a number of fitness-related tasks in mammals, from foraging and danger avoidance, to the complex behavioural processes of individual recognition, mate choice and maternal care (Firestein 2001, Brennan & Kendrick 2006). In this context, olfactory receptor genes are constantly under pressure to ensure the detection of olfactants that are fundamental for a species' success. In order to ensure the ability to distinguish novel important odorants in the environment, variability at OR loci is selected for via positive selection. OR variation is thus generated via mutation and gene duplication and, if it fails to translate into functionally advantageous variants, pseudogenisation is likely to occur. Additionally, the importance of detecting fundamental odorants that are common to all species and habitats, generates a strong opposing force, favouring the maintenance of functionally unaltered OR genes across species and in time. Trans-species polymorphisms are indicative of this phenomenon, accumulating under strong purifying selective forces.

Thus, all the classic aspects of 'birth and death' evolution which characterise the vertebrate OR repertoire, are suggestive of the incredible responsiveness of the olfactory system, and are evident across the bathyergid OR7 subgenome: from adaptive evolution, to the presence of an array of OR polymorphisms (generated by

University of Cape Town

adaptive forces) and ultimately pseudogenisation. Furthermore, the importance of purifying selection emerges as an important addition to the 'birth and death' model of evolution in the OR multigene family. Based on these findings, and depending on the biological processes in which they are involved, bathyergid OR7 genes are likely to be responsible for the recognition of a broad variety of odorant chemicals, ranging from fundamental odorants (presumably detected by ORs under purifying selection - Clade D, Chapter 3), to classes of odorants that may require a more 'flexible' detection strategy (subject to adaptive evolution - clades A and C, Chapter 3). Importantly, this finding is in contrast to the commonly accepted idea that ORs belonging to the same gene family necessarily have similar functions (Malnic et al. 2004), and further supports the model of OR classification based on 'process' proposed in Chapter 3.

Paradoxically, in a time of sophisticated whole-genome studies that allow for the complete characterisation of OR subgenomes in the model organisms studied (e.g. Niimura and Nei 2003, 2005a, 2005b, 2007, Niimura 2009), both *in vitro* and *in vivo* approaches for the functional characterisation of olfactory receptors and their range of ligands remain a major challenge (Saito et al. 2009). In this context, the OR classification approach based on 'process', as proposed in this thesis, complements the traditional approach based on 'pattern', and can enlighten our understanding of the evolution of this large, complicated gene family.

Undoubtedly, the recently published *H. glaber* genome (*Naked Mole-Rat Genome Resource*, http://naked-mole-rat.org, July 2011) represents an exciting opportunity to complement the findings from the comparative genetic study presented in this thesis; intensive data-mining within the naked mole-rat genome will enable us to characterise the entire OR repertoire of *H. glaber*. An estimate of the magnitude of this OR subgenome will enable us to estimate to what degree the OR7 repertoire, as characterised in this study across the Bathyergidae, is truly representative of OR7 genomic diversity. Furthermore, the presence of ORs from other gene families can be verified and this will enable future studies to target specific portions of the OR subgenome across bathyergids. Typically, this would help to identify expanded OR families which may have responded to the selective regime of the subterranean niche (testing the Ecogroup Hypothesis, Chapter 4). The availability of a whole-genome

University of Cape Town

sequence for this species is also crucial for the future design of Bathyergidae-specific oligonucleotide primers, allowing for the targeted analysis of distinct OR genes.

It is important to remember, however, that the genetic analysis of OR subgenomes presents two unavoidable limitations, which can only be remedied with complementary OR expression studies. The first limitation comes from the finding that a number of ORs traditionally classified as 'pseudogenes', are in fact transcribed in humans (Zhang et al. 2007b) and this is suggestive of a role for OR pseudogenes in regulating gene expression and mediating atypical OR functions. If a number of OR 'pseudogenes' are indeed functional, the ratio of OR genes:pseudogenes commonly used to define the olfactory ability of a species, and the primary tool in comparative studies on the role of the environment (e.g. Kishida et al. 2007, Hayden et al. 2010, or the Ecogroup Hypothesis in this thesis), is both biased and potentially misleading. The second limitation of genetic OR studies is that of the unknown function of 'ectopically-expressed' OR genes, which appear to be numerous and have been suggested to carry out non-olfactory functions outside the olfactory system (Spehr et al. 2003, De la Cruz et al. 2009). Thus, OR genes classified as 'functional' based on classic DNA sequence features may in fact bias the estimation of functional OR repertoire size and diversity when including those OR genes expressed outside the olfactory system.

To fully understand the complex mechanisms of olfaction in African mole-rats it would be interesting to complement this research on ORs with the study of the vomeronasal olfactory system in Bathyergidae. In the naked mole-rat, the vomeronasal organ, VNO, is degenerate and thus appears to be non-functional (Smith et al. 2007). If this is a common trait in all bathyergid species, it would be interesting to investigate whether vomeronasal receptors, VRs, which are generally numerous in rodents (Figure 1.5, Chapter 1), are expressed in another portion of the olfactory system. VRs are fundamental to pheromonal communication in rodents; candidate ligands for VRs, particularly V2Rs, include a number of non-volatile pheromones, such as major urinary proteins (MUPs) (Krieger et al. 1999, Chamero et al. 2007), major histocompatibility complex (MHC) peptides (Leinders-Zufall et al. 2004, He et al. 2008) and the exocrine gland-secreting peptide (ESP) family (Kimoto et al. 2005, 2007). It is also possible that in the absence of a functional VNO, pheromonal

University of Cape Town

communication may be mediated by ORs expressed in the main olfactory epithelium, as has been suggested for humans (Wang et al. 2007). Assuming that, like *H. glaber*, all bathyergids lack a functional VNO, pheromonal communication normally conveyed by the rodent VNO may either occur via VRs expressed outside the VNO, or via ORs in the MOE, or via a combination of the two 'compensating' mechanisms.

Although the results presented in this thesis are not directly consistent with a putative role for OR7 bathyergid genes in social communication (Chapter 4), they certainly highlight the importance for investigating a role for ORs in pheromonal signalling. Exploring this aspect of olfactory dynamics in the Bathyergidae would undoubtedly help us gain a more thorough understanding of the complex role of olfaction in African mole-rat physiology and evolution.

University of Cape Town

References

Abe H, Kanaya S, Komukai T, Takahashi Y, Sasaki S (1990) Systematization of semantic descriptions of odors. *Analytica Chimica Acta,* **239**: 73-85.

Ache BW & Young JM (2005) Olfaction: diverse Species, conserved principles. *Neuron*, **48**: 417–430.

Adkins RM, Gelke EL, Rowe D, Honeycutt RL (2001) Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Molecular Biology and Evolution*, **18**: 777–791.

Aeschlimann PB, Häberli MA, Reusch TBH, Boehm T, Milinski M (2003) Female sticklebacks *Gasterosteus aculaetus* use self-reference to optimize MHC allele number during mate selection. *Behavioural Ecology and Sociobiology,* **54**: 119–126.

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 3490–3494.

Alioto TS and Ngai J (2005) The odorant receptor repertoire of teleost fish. *BMC Genomics,* **6**: 173.

Al-Ubaidi MR, Pittler SJ, Champagne MS, Triantafyllos JT, McGinnis JF, Baehr W (1990) Mouse Opsin. Gene structure and molecular basis of multiple transcripts. *The Journal of Biological Chemistry,* **265**: 20563-20569.

Axelson JF, Smith M, Duarte M (1999) Prenatal flutamide treatment eliminates the adult male rat's dependency upon vasopressin when forming social-olfactory memories. *Hormones and Behavior*, **36**: 109–18.

Balakirev ES and Ayala FJ (2003) Pseudogenes: are they "junk" or functional DNA? *Annual Review of Genetics,* **37**: 123–151.

Bardeleben C, Moore RL, Wayne RK (2005) A molecular phylogeny of the Canidae based on six nuclear loci. *Molecular Phylogenetics and Evolution*, **37:** 815–831.

Bargmann CI (2006) Comparative chemosensation from receptors to ecology. *Nature,* **444:** 295–301.

Batra SWT (1966) Nests and social behavior of halictine bees of India. *Indian Journal of Entomology*, **28**: 375-393.

Baxi KN, Dorries KM, Eisthen HL (2006) Is the vomeronasal system really specialized for detecting pheromones? *Trends in Neurosciences,* **29**: 1–7.

Bellringer JF, Pratt HPM, Keverne EB (1980) Involvement of the vomeronasal organ and prolactin in pheromonal induction of delayed implantation in mice. *Journal of Reproduction and Fertility,* **59**: 223-228.

University of Cape Town

Ben-Arie N, Lancet D, Taylor C, Khen M, Walker N, Ledbetter DH, Carrozzo R, Patel K, Sheer D, Lehrach H, *et al.* (1994) Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. *Human Molecular Genetics*, **3:** 229-235.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**:289-300.

Bennett NC (1988) *The trend towards sociality in three species of southern African mole-rats (Bathyergidae): causes and consequences.* PhD thesis (unpublished), University of Cape Town, South Africa.

Bennett NC (1989) The social structure and reproductive biology of the common mole-rat, *Cryptomys hottentotus hottentotus* and remarks on the trends in reproduction and sociality in the family Bathyergidae. *Journal of Zoology, London,* **219:** 45–59.

Bennett NC (1992) Aspects of social behaviour in a captive colony of the common mole-rat, *Cryptomys hottentotus hottentotus*. *Zeitschrift für Säugetierkunde*, **57:** 294–309.

Bennett NC and Faulkes CG (2000) *African mole-rats ecology and eusociality*. Cambridge University Press, Cambridge, UK.

Bensafi M, Brown WM, Khan R, Levenson B, Sobel N (2004) Sniffing human sex-steroid derived compounds modulates mood, memory and autonomic nervous system function in specific behavioral contexts. *Behavioural Brain Research,* **152**:11–22.

Benton R, Sachse S, Michnick SW, Vosshall LB (2006) Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS Biology,* **4:** e20.

Berghard A and Buck LB (1996) Sensory transduction in vomeronasal neurons: evidence for Gαo, Gαi2, and adenylyl cyclase II as major components of a pheromone signaling cascade. *Journal of Neuroscience,* **16**:909–918.

Betts MJ and Russell MB (2003) Amino acid properties and consequences of substitution. In *Bioinformatics for geneticists.* Barnes MR and Gray IC (eds), Wiley.

Bhatnagar KP and Smith TD (2010) The human vomeronasal organ. Part VI: A nonchemosensory vestige in the context of major variations of the mammalian vomeronasal organ. *Current Neurobiology,* **1**: 1-9.

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL and Purvis A (2007) The delayed rise of present-day mammals. *Nature,* **446**: 507-512.

Bishop JM, Jarvis JUM, Spinks AC, Bennett NC, O'Ryan C (2004). Molecular insight into patterns of colony composition and paternity in the common mole-rat *Cryptomys hottentotus hottentotus*. *Molecular Ecology,* **13**: 1217–1229.

Bishop JM, O'Ryan C, Jarvis JUM (2007) Social common mole-rats enhance outbreeding via extra-pair mating. *Biology Letters,* **3**: 176–179.

University of Cape Town

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW, Huchon D (2009) Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology,* **9:** 71.

Bluthe RM and Dantzer R (1990) Social recognition does not involve vasopressinergic neurotransmission in female rats. *Brain Research*, **535**: 301–304.

Bluthe RM and Dantzer R (1992) Chronic intracerebral infusion of vassopressin antagonist modulates social recognition in rat. *Brain Research*, **572**: 261–264.

Bonneaud C, Chastel O, Federici P, Westerdahl H, Sorci G (2006) Complex Mhc-based mate choice in a wild passerine. *Proceedings of the Royal Society B,* **273**: 1111–1116.

Borowsky B, Adham N, Jones KA, Raddatz R, Artymyshyn R, Ogozalek KL, Durkin MM, Lakhlani PP, Bonini JA, Pathirana S, *et al.* (2001) Trace amines: identification of a family of mammalian G protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America,* **98**: 8966–8971.

Boughman JW (2002) How sensory drive can promote speciation. *Trends in Ecology and Evolution,* **17**: 571–577.

Boulay F, Tardif M, Brouchon L, Vignais P (1990) The human N-formylpeptide receptor. Characterization of two cDNA isolates and evidence for a new subfamily of G-protein-coupled receptors. *Biochemistry,* **29**: 11123–11133.

Branscomb A, Seger J, White RL (2000) Evolution of odorant receptors expressed in mammalian testes. *Genetics,* **156**: 785–797.

Braude SH and Ciszek D (1998) Survival of naked mole-rats marked by implantable transponders and toe-clipping. *Journal of Mammalogy,* **79**: 360-363.

Breer H, Fleischer J, Strotmann J (2006) The sense of smell: multiple olfactory subsystems. *Cellular and Molecular Life Sciences,* **63:** 1465–1475.

Brennan PA and Kendrick KM (2006) Mammalian social odours: attraction and individual recognition. *Philosophical Transactions of the Royal Society B,* **361**: 2061–2078.

Brennan PA and Zufall F (2006) Pheromonal communication in vertebrates. *Nature*, **444:** 308–315.

Brett RA (1991) The population structure of naked mole-rat colonies. In: *The Biology of the Naked Mole-Rat*. Sherman PW, Jarvis JUM, Alexander RD (eds), Princeton University Press, Princeton, NJ, pp 97-136.

Bridges TM and Lindsley CW (2008) G-protein-coupled receptors: from classical modes of modulation to allosteric mechanisms. *ACS Chemical Biology,* **3:** 530–541.

Brown WA and G Heninger G (1975) Cortisol, growth hormone, free fatty acids, and experimentally evoked affective arousal. *American Journal of Psychiatry,* **132**: 1172-1176.

Buck L and Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**: 175-187.

University of Cape Town

Buffenstein R and Yahav S (1991a) Cholecalciferol has no effect on calcium and inorganic phosphorus balance in a naturally cholecalciferol-deplete subterranean mammal, the naked mole rat (*Heterocephalus glaber*). *Journal of Endocrinology,* **129**: 21-26.

Buffenstein R and Yahav S (1991b) Is the naked mole-rat *Hererocephalus glaber* an endothermic yet poikilothermic mammal? *Journal of Thermal Biology,* **16**: 227-232.

Burda H (1995) Individual recognition and incest avoidance in eusocial common mole-rats rather than reproductive suppression by parents. *Experientia,* **51**: 411 413.

Burda H (1997) Evolution of life history strategies in African mole-rats (Bathyergidae) (abstract). *7th International Theriological Congress, Acapulco, Mexico, 6-11 September*, p 24.

Burda H (1999) Syndrome of eusociality in African subterranean mole-rats (Bathyergidae, Rodentia), its diagnosis and aetiology. In *Evolutionary theory and processes: modern perspectives*. Wasser SP (ed), Kluwer, Dordrecht, Netherlands, pp 385-418.

Burda H, Honeycutt RL, Begall S, Locker-Grütjen O, Scharff A (2000) Are naked and common mole-rats eusocial and if so, why? *Behavioural Ecology and Sociobiology,* **47**: 293–303.

Burland TM, Bennett NC, Jarvis JUM, Faulkes CG (2002) Eusociality in African mole-rats: new insights from patterns of genetic relatedness in the Damaraland mole-rat (*Cryptomys damarensis*). *Proceedings of the Royal society of London B,* **269**: 1025–1030.

Burland TM, Bennett NC, Jarvis JUM, Faulkes CG (2004) Colony structure and parentage in wild colonies of cooperatively breeding Damaraland mole-rats suggest incest avoidance alone may not maintain reproductive skew. *Molecular Ecology,* **13**: 2371–2379.

Caldeira K, Wickett ME (2005) Ocean model predictions of chemistry changes from carbon dioxide emissions to the atmosphere and ocean. *Journal of Geophysical Research,* **110**: C09S04.

Campbell A (2008) Attachment, aggression and affiliation: the role of oxytocin in female social behavior. *Biological Psychology,* **77**: 1–10.

Castillo AH, Cortinas MN, Lessa EP (2005) Rapid diversification of South American tuco-tucos (Ctenomys; Rodentia, Ctenomidae): contrasting mitochondrial and nuclear intron sequences. *Journal of Mammalogy*, **86**:170–179.

Chamero P, Marton TF, Logan DW, Flanagan K, Cruz JR, Saghatelian A, Cravatt BF, Stowers L (2007) Identification of protein pheromones that promote aggressive behaviour. *Nature,* **450**: 899–902.

Charles-Dominique P (1977b) Urine marking and territoriality in *Galago alleni*. A field study by radio telemetry. *Zeitschrift für Tierpsychologie,* **43**: 113-138.

Chatterjee HJ, Ho SY, Barnes I, Groves C (2009) Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evolutionary Biology*, **9**: 259.

Chess A, Simon I, Cedar H, Axel R (1994) Allelic inactivation regulates olfactory receptor gene expression. *Cell,* **78**: 823–834.

University of Cape Town

Clarke FM and Faulkes CG (1997) Dominance and queen succession in captive colonies of the eusocial naked mole-rat, *Heterocephalus glaber*. *Proceedings of the Royal society of London B,* **264**: 993–1000.

Colbourne JK, Wilson CC, Hebert PDN (2006) The systematics of Australian Daphnia and Daphniopsis (Crustacea: Cladocera): a shared phylogenetic history transformed by habitat-specific rates of evolution. *Biological Journal of the Linnean Society*, **89**: 469–488.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics,* **21**: 3674-3676.

Costa JT and Fitzgerald TD (2005) Social terminology revisited: Where are we ten years later? *Annales Zoologici Fennici*, **42**: 559-564.

Crespi BJ and Yanega D (1995) The definition of eusociality. *Behavioral Ecology*, **6**: 109-115.

Crish SD, Rice FL, Park TJ, Comer CM (2003) Somatosensory organization and behavior in naked mole-rats I: vibrissa-like body hairs comprise a sensory array that mediates orientation to tactile stimuli. *Brain Behavior and Evolution,* **62**: 141–151.

Dagert M and Ehrlich SD (1979) Prolonged incubation in calcium chloride improves competence of *Escherichia coli* cells. *Gene,* **6:** 23-28.

Dantzer R, Tazi A, Bluthe RM (1990) Cerebral lateralization of olfactory-mediated affective processes in rats. *Behavioural Brain Research*, **40**: 53–60.

Darden TR (1972) Respiratory adaptations of a fossorial mammal, the pocket gopher (*Thomomys bottae*). *Journal of Comparative Physiology,* **78**: 121-137.

Davies KC and Jarvis JUM (1986) The burrow systems and burrowing dynamics of the mole-rats *Bathyergus suillus* and *Cryptomys hottentotus* in the fynbos of the south-western Cape, South Africa. *Journal of Zoology, London,* **209**: 125-147.

De Graaf G (1981) *The rodents of Southern Africa*, *Butterworth, Johannesburg*.

De la Cruz O, Blekhman R, Zhang X, Nicolae D, Firestein S, Gilad Y (2009) A signature of evolutionary constraint on a subset of ectopically expressed olfactory receptor genes. *Molecular Biology and Evolution,* **26**: 491–494.

Del Punta K, Leinders-Zufall T, Rodriguez I, Jukam D, Wysocki C, Ogawa S, Zufall F, Mombaerts P (2002) Deficient pheromone responses in mice lacking a cluster of vomeronasal receptor genes. *Nature,* **419**: 70–74.

Dixson AF (1998) *Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Human Beings*. Oxford University Press, Oxford.

Dixson DL, Munday PL, Jones GP (2010) Ocean acidification disrupts the innate ability of fish to detect predator olfactory cues. *Ecology Letters,* **13**: 68–75.

Doherty PC and Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, **256:** 50–52.

University of Cape Town

Donaldson ZR and Young LJ (2008) Oxytocin, vasopressin, and the neurogenetics of sociality. *Science,* **322**: 900 –904.

Doron-Faigenboim A, Stern A., Mayrose I., Bacharach E., and Pupko T (2005) Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics,* **21**: 2101-2103.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics,* **36**:1326–1329.

Dulac C and Axel R (1995) A novel family of genes encoding putative pheromone receptors in mammals. *Cell,* **83**: 195–206.

Duret L, Chureau C, Samain S, Weissenbach J, Avner P (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science,* **312**: 1653–1655.

Duret L and Mouchiroud D (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Molecular Biology and Evolution,* **17**: 68–74.

Ebensperger LA and Blumstein DT (2006) Sociality in New World hystricognath rodents is linked to predators and burrow digging. *Behavioral Ecology*, **17**: 410–418.

Eccles R (1982) Autonomic innervation of the vomeronasal organ of the cat. *Physiology & Behavior*, **28:** 1011-1015.

Ekblom R, Sæther SA, Fiske P, Kålås JA and Höglund J (2010) Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. *Genetica,* **138**: 453–461.

Eloff G (1958) The structural and functional degeneration of the eye of South African rodent moles *Cryptomys bigalkei* and *Bathyergus maritimus*. *South African Journal of Science,* **54**: 293-302.

Emes RD, Beatson SA, Ponting CP, Goodstadt L (2004) Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. *Genome Research*, **14:** 591-602.

Fabry VJ, Seibel BA, Feely RA, Orr JC (2008) Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES Journal of Marine Science,* **65**: 414–432.

Faulkes CG (1990) Social suppression of reproduction in the naked mole-rat *Heterocephalus glaber*. *PhD thesis (unpublished), University of London, United Kingdom*.

Faulkes CG, Abbott DH, Jarvis JUM (1990) Social suppression of ovarian cyclicity in captive and wild colonies of naked mole-rats, *Heterocephalus glaber*. *Journal of Reproduction and Fertility,* **88**: 559-568.

Faulkes CG and Abbott DH (1993) Evidence that primer pheromones do not cause social suppression of reproduction in male and female naked mole-rats (*Heterocephalus glaber*). *Journal of Reproduction and Fertility,* **99**: 225-230.

Faulkes CG, Bennett NC, Bruford MW, O'Brien HP, Aguilar GH, Jarvis JUM (1997) Ecological constraints drive social evolution in the African mole-rats. *Proceedings of the Royal society of London B,* **264**: 1619-1627.

University of Cape Town

Faulkes CG (1998) Naked mole-rats. *In: Encyclopedia of reproduction, Knobil E, Neill JD (eds), Academic Press, San Diego,* **3:** 321-326.

Faulkes CG, Verheyen W, Jarvis JUM, Bennett NC (2004) Phylogeographic patterns of speciation and geneic divergence in African mole-rats (Family Bathyergidae). *Molecular Ecology*, **13**: 613-629.

Feldmesser E, Olender T, Khen M, Yanai I, Ophir R, Lancet D (2006) Widespread ectopic expression of olfactory receptor genes. *BMC Genomics*, **7**: 121.

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution,* **17**: 368–376.

Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution,* **39**: 783-791.

Ferrero DM and Liberles SD (2010) The secret codes of mammalian scents. *WIREs Systems Biology and Medicine*, **2:** 23-33.

Figueroa F, Gunther E, Klein J (1988) MHC polymorphism predating speciation. *Nature,* **335**: 265-267.

Finn S and Civetta A (2010) Sexual selection and the molecular evolution of ADAM proteins. *Journal of Molecular Evolution*, **71**: 231–240.

Firestein S (2001) How the olfactory system makes sense of scents. *Nature,* **413**: 211-218.

Fisher HS, Swaisgood RR, Fitch Snyder H, (2003) Counter marking by male pygmy lorises (*Nycticebus pygmaeus*): do females use odour cues to select mates with high competitive ability? *Behavioural Ecology and Sociobiology,* **53:** 123-130.

Fleischer J, Schwarzenbacher K, Besser S, Hass N, Breer H (2006) Olfactory receptors and signalling elements in the Grueneberg ganglion. *Journal of Neurochemistry,* **98:** 543–554.

Fleischer J, Schwarzenbacher K, Breer H (2007) Expression of trace amine-associated receptors in the Grueneberg ganglion. *Chemical Senses,* **32**: 623–631.

Fleischer J, Breer H and Strotmann J (2009) Mammalian Olfactory Receptors. *Frontiers in Cellular Neuroscience,* **3**: 1-10.

Foster KR and Ratnieks FLW (2005) A new social vertebrate? *Trends in Ecology and Evolution,* **20:** 363–364.

Freitag J, Krieger J, Strotmann J, Breer H (1995) Two classes of olfactory receptors in *Xenopus laevis*. *Neuron,* **15**: 1383-1392.

Freitag J, Ludwig G, Andreini I, Rossler P, Breer H (1998) Olfactory receptors in aquatic and terrestrial vertebrates. *Journal of Comparative Physiology A,* **183**: 635-650.

Fu YX and Li WH (1993) Statistical tests of neutrality of mutations. *Genetics,* **133**: 693-709.

Fukuda N and Touhara K (2006) Developmental expression patterns of testicular olfactory receptor genes during mouse spermatogenesis. *Genes to Cells,* **11:** 71–81.

University of Cape Town

Fülle HJ, Vassar R, Foster DC, Yang RB, Axel R, Garbers DL (1995) A receptor guanylyl cyclase expressed specifically in olfactory sensory neuron. *Proc. Natl. Acad. Sci. U.S.A.* **92:** 3571–3575.

Fuss SH, Omura M, Mombaerts P (2007) Local and *cis* effects of the H element on expression of odorant receptor genes in mouse. *Cell,* **130**: 373–384.

Gaillard I, Rouquier S, Pin JP, Mollard P, Richerd S, Barnabé C, Demaille J and Giorgi D (2002) A single olfactory receptor specifically binds a set of odorant molecules. *European Journal of Neuroscience,* **15**: 409–418.

Garrigan D and Hendrick PN (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics,* **53**: 483-489.

Gether U (2000) Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocrine Reviews,* **21**: 90–113.

Ginzel MD (2010) Olfactory signals. In *Encyclopedia of Animal Behaviour.* Breed M and Moore J (eds), Elsevier Ltd., Oxford, pp 584–588.

Gilad Y, Bustamante CD, Lancet D and Paabo S (2003) Natural selection on the olfactory receptor gene family in humans and chimpanzees. *American Journal of Human Genetics,* **73**: 489–501.

Gilad Y and Lancet D (2003) Population differences in the human functional olfactory repertoire. *Molecular Biology and Evolution*, **20:** 307-314.

Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biology,* **2**: e5.

Glusman G, Clifton S, Roe B, Lancet D (1996) Sequence analysis in the olfactory receptor gene cluster on human chromosome 17: recombinatorial events affecting receptor diversity. *Genomics,* **37:** 147-160.

Glusman G, Bahar A, Sharon D, Pilpel Y, White J, Lancet D (2000a) The olfactory receptor gene superfamily: Data mining, classification, and nomenclature. *Mammalian Genome,* **11**: 1016-1023.

Glusman G, Sosinsky A, Ben-Asher E, Avidan N, Sonkin D, Bahar A, Rosenthal A, Clifton S, Roe B, Ferraz C, *et al.* (2000b) Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. *Genomics*, **63:** 227-245.

Glusman G, Yanai I, Rubin I, Lancet D (2001) The complete human olfactory subgenome. *Genome Research,* **11**: 685-702.

Godfrey PA, Malnic B, Buck LB (2004) The mouse olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America,* **101**: 2156–2161.

Gosling ML, Roberts SC, Thornton EA, Andrew MJ (2000) Life history costs of olfactory status signalling in mice. *Behavioral Ecology and Sociobiology,* **48:** 328-332.

University of Cape Town

Götz S, García-Gómez JM, Terol J, Williams TD, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research,* **36**: 3420–3435.

Graves BM and Halpern M (1990) Roles of vomeronasal organ chemoreception in tongue flicking, exploratory and feeding behaviour of the lizard, *Chalcides ocellatus*. *Animal Behaviour,* **39**: 692-698.

Griff IC and Reed RR (1995) The genetic basis for specific anosmia to isovaleric acid in the mouse. *Cell*, **83:** 407-414.

Griffiths AJ, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (2000) *An Introduction to Genetic Analysis*. Seventh Edition © W.H. Freeman and Company, New York.

Gross-Isseroff R, Ophir D, Bartana A, Voet H, Lancet D (1992) Evidence for genetic determination in human twins of olfactory thresholds for a standard odorant. *Neuroscience Letters,* **141:** 115-118.

Grosser BI, Monti-Bloch L, Jennings-White C, Berliner DL (2000) Behavioral and electrophysiological effects of androstadienone, a human pheromone. *Psychoneuroendocrinology,* **25**: 289 –299.

Grus WE, Shi P, Zhang YP, Zhang J (2005) Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals. *Proceedings of the National Academy of Sciences of the United States of America,* **102**: 5767–5772.

Grus WE, Shi P, Zhang J (2007) Largest vertebrate vomeronasal type 1 receptor gene repertoire in the semiaquatic platypus. *Molecular Biology and Evolution,* **24**: 2153–2157.

Haddad R, Khan R, Takahashi YK, Mori K, Harel D, Sobel N (2008) A metric for odorant comparison. *Nature Methods,* **5**: 425–29.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS (2007) Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics*, **177**: 1321–1335.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* **41**: 95-98.

Halpern M and Kubie JL (1980) Chemical access to the vomeronasal organs of garter snakes. *Physiology & Behavior*, **24**: 367-371.

Halpern M and Martinez-Marcos A (2003) Structure and function of the vomeronasal system: an update. *Progress in Neurobiology,* **70**: 245–318.

Hamilton LD, Rellini AH, Meston CM (2008) Cortisol, sexual arousal, and affect in response to sexual stimuli. *Journal of Sexual Medicine,* **5**: 2111–2118.

Hayden S, Bekaert M, Crider TA, Marini S, Murphy WJ, Teeling EC (2010) Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Research,* **20**: 1-9.

University of Cape Town

He J, Ma L, Kim S, Nakai J, Yu CR (2008) Encoding gender and individual information in the mouse vomeronasal organ. *Science,* **320**: 535–538.

Hebert PDN, Remigio EA, Colbourne JK, Taylor DJ, Wilson CC (2002) Accelerated molecular evolution in halophilic crustaceans. *Evolution*, **56**: 909–926.

Hedrick PW (1999) Balancing selection and MHC. *Genetica,* **104:** 207–214.

Heffner RS and Heffner HE (1993) Degenerate hearing and sound localization in naked mole rats (*Heterocephalus glaber*), with an overview of central auditory structures. *Journal of Comparative Neurology,* **331**: 418-433.

Herlyn H and Zischler H (2007) Sequence evolution of the sperm ligand zonadhesin correlates negatively with body weight dimorphism in primates. *Evolution*, **61:** 289-298.

Herrada G and Dulac C (1997) A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell,* **90**: 763–773.

Heth G, Todrank J, Begall S, Koch R, Zilbiger Y, Nevo E, Braude SH, Burda H (2002a) Odours underground: subterranean rodents may not forage "blindly". *Behavioral Ecology and Sociobiology,* **52**: 53-58.

Heth G, Todrank J and Burda H (2002b) Individual odor similarities within colonies and across species of *Cryptomys* mole rats. *Journal of Mammalogy*, **83**: 569-575.

Heth G, Todrank J, Begall S, Wegner RE and Burda H (2004) Genetic relatedness discrimination in eusocial *Cryptomys anselli* mole-rats, Bathyergidae, Rodentia. *Folia Zoologica,* **53**: 269-278.

Holloway AK and Begun DJ (2004) Molecular evolution and population genetics of duplicated accessory gland protein genes in *Drosophila*. *Molecular Biology and Evolution*, **21**: 1625–1628.

Holmes WG and Sherman PW (1983) Kin recognition in animals. *American Scientist,* **71**: 46-55.

Holmes MM, Goldman BD, Forger NG (2008) Social status and sex independently influence androgen receptor expression in the eusocial naked mole-rat brain, *Hormones and Behavior,* **54**: 278–285.

Honeycutt RL, Allard MW, Edwards SV, Schlitter DA (1991) Systematics and evolution of the family Bathyergidae. In *The Biology of the Naked Mole-Rat.* Sherman PW, Jarvis JUM, Alexaner RD (eds), Princeton University Press, Princeton, NJ, pp 45-65.

Huchon D and Douzery EJ (2001) From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. *Molecular Phylogenetics and Evolution,* **20:** 238–251.

Hughes AL and Nei M (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Molecular Biology and Evolution,* **6**: 559–579.

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity,* **99**: 364-373.

University of Cape Town

Hughes NK, Kelley JL, Banks PB (2009) Receiving behaviour is sensitive to risks from eavesdropping predators. *Oecologia,* **160**: 609–617.

Hughes NK, Price CJ, Banks PB (2010a) Predators are attracted to the olfactory signals of prey. *PLoS ONE*, **5**: e13114.

Hughes NK, Korpimäki E, Banks PB (2010b) The predation risks of interspecific eavesdropping: weasel-vole interactions. *Oikos,* **119**: 1210–1216.

Hurst JL and Beynon RJ (2004) Scent wars: the chemobiology of competitive signalling in mice. *Bioessays,* **26**: 1288–1298.

Ingram CM, Burda H, Honeycutt RL (2004) Molecular phylogenetics and taxonomy of the African mole-rats, genus *Cryptomys* and the new genus *Coetomys Gray*, 1864. *Molecular Phylogenetics and Evolution,* **31**: 997-1014.

Insel TR and Young LJ (2000) Neuropeptides and the evolution of social behavior. *Current Opinion in Neurobiology,* **10**: 784–789.

Ishii T, Serizawa S, Kohda A, Nakatani H, Shiroishi T, Okumura K, Iwakura Y, Nagawa F, Tsuboi A, Sakano H (2001) Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurons. *Genes to Cells,* **6:** 71–78.

Ishii T, Hirota J, Mombaerts P (2003) Combinatorial coexpression of neural and immune multigene families in mouse vomeronasal sensory neurons. *Current Biology,* **13:** 394–400.

Ishii T and Mombaerts P (2008) Expression of nonclassical class I major histocompatibility genes defines a tripartite organization of the mouse vomeronasal system. *Journal of Neuroscience,* **28**: 2332–2341.

Iwema CL, Fang H, Kurtz DB, Youngentob SL, Schwob JE (2004) Odorant receptor expression patterns are restored in lesion-recovered rat olfactory epithelium. *Journal of Neuroscience,* **24:** 356–369.

Jacob S, Kinnunen LH, Metz J, Cooper M, McClintock MK (2001a) Sustained human chemosignal unconsciously alters brain function. *Neuroreport,* **12**: 2391–2394.

Jacob S, Hayreh DJ, McClintock MK (2001b) Context-dependent effects of steroid chemosignals on human physiology and mood. *Physiology & Behavior,* **74**: 15–27.

Jacob S, McClintock MK, Zelano B, Ober C (2002) Paternally inherited HLA alleles are associated with women's choice of male odour. *Nature Genetics*, **30:** 175–179.

Jacoby E, Bouhelal R, Gerspacher M, Seuwen K (2006) The 7 TM G-protein-coupled receptor target family. *Chemmedchem,* **1:** 761–782.

Jakupovic J, Kang N, Baum MJ (2008) Effect of bilateral accessory olfactory bulb lesions on volatile urinary odor discrimination and investigation as well as mating behavior in male mice. *Physiology & Behavior*, **93:** 467–473.

Jarvis JUM (1981) Eu-sociality in a mammal: cooperative breeding in naked mole-rat *Heterocephalus glaber* colonies. *Science,* **212**: 571-573.

Jarvis JUM and Bennett NC (1990) The evolutionary history, population biology and social structure of African mole-rats: family Bathyergidae. In *Evolution of Subterranean Mammals*

University of Cape Town

at the Organismal and Molecular Llevels. Nevo E and Reig OA (eds), Liss, New York, pp 97-128.

Jarvis JUM (1991) Reproduction of naked mole-rats. In *The Biology of the Naked Mole-Rat.* Sherman PW, Jarvis JUM, Alexander RD (eds), Princeton University Press, Princeton, NJ, pp 384-425.

Jarvis JUM and Bennett NC (1991) Ecology and behavior of the family Bathyergidae. In *The biology of the Naked Mole-Rat.* Sherman PW, Jarvis JUM, Alexander RD (eds), Princeton University Press, Princeton, NJ, pp 66-96.

Jarvis JUM and Bennett NC (1993) Eusociality has evolved independently in two genera of bathyergid mole-rats – but occurs in no other subterranean mammal. *Behavioral Ecology and Sociobiology,* **33**: 353-360.

Jarvis JUM, O'Riain MJ, Bennett NC, Sherman PW (1994) Mammalian eusociality: a family affair. *Trends in Ecology and Evolution,* **9**: 47-51.

Jarvis JUM, Bennett NC, Spinks AC (1998) Food availability and foraging by wild colonies of Damaraland mole-rats (*Cryptomys damarensis*): implications for sociality. *Oecologia,* **113**: 290-298.

Jarvis JUM and Sherman PW (2002) Heterocephalus glaber. *Mammalian Species*, **706:** 1–9.

Jeltema MA and Southwick EW (1986) Evaluations and application of odor profiling. *Journal of Sensory Studies,* **1**: 123-136.

Jennions MD and Petrie M (2000) Why do females mate multiply? A review of the genetic benefits. *Biological Reviews*, **75:** 21–64.

Jolliffe (1986) *Principal Component Analysis*. Springer-Verlag, New York.

Jones DT and Reed RR (1989) An olfactory neuron specific G-protein involved in odorant signal transduction. *Science,* **244**: 790–95.

Judd TM and Sherman PW (1996) Naked mole-rats recruite colony mates to food sources. *Animal Behavior,* **52:** 957-969.

Juilfs DM, Fülle HJ, Zhao AZ, Houslay MD, Garbers DL, Beavo JA (1997) A subset of olfactory neurons that selectively express cGMP-stimulated phosphodiesterase (PDE2) and guanylyl cyclase-D define a unique olfactory signal transduction pathway. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 3388–3395.

Kajiya S, Hirokawa T, Oka Y, Suwa M, Touhara K (2001) Molecular bases of odor discrimination: reconstitution of olfactory receptors that recognize overlapping sets of odorants. *Journal of Neuroscience,* **21**: 6018-6025.

Kalamatianos T, Faulkes C, Oosthuizen M, Poorun R, Bennett N, Coen CW (2010) Telencephalic binding sites for oxytocin and social organization: a comparative study of eusocial naked molerats and solitary cape mole-rats. *Journal of Comparative Neurology,* **518**: 1792-1813.

Kaluza JF, Gussing F, Bohm S, Breer H, Strotmann J (2004) Olfactory receptors in the mouse septal organ. *Journal of Neuroscience Research,* **76:** 442–452.

University of Cape Town

Kambere MB and Lane RP (2007) Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. *BMC Neuroscience,* **8**: S2. doi: 10.1186/ 1471-2202-8-S3-S2.

Kappeler PM (1998) To whom it may concern: the transmission and function of chemical signals in *Lemur catta*. *Behavioral Ecology and Sociobiology,* **42**: 411-421.

Karlson P and Luscher M (1959) Pheromones': a new term for a class of biologically active substances. *Nature,* **183**: 55–56.

Katada S, Hirokawa T, Oka Y, Suwa M, Touhara K (2005) Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. *The Journal of Neuroscience*, **25**: 1806 –1815.

Kato A, Katada S, Touhara K (2008) Amino acids involved in conformational dynamics and G protein coupling of an odorant receptor: targeting gain-of-function mutation. *Journal of Neurochemistry,* **107:** 1261–1270.

Keller L and Perrin N (1995) Quantifying the level of eusociality. *Proceedings of the Royal society of London B*, **260**: 311-315.

Keller A, Zhuang H, Chi Q, Vosshall LB, Matsunami H (2007) Genetic variation in a human odorant receptor alters odour perception. *Nature,* **449**: 468-472.

Keller A and Vosshall LB (2008) Better smelling through genetics: mammalian odor perception. *Current Opinion in Neurobiology,* **18**: 364–369.

Kendrick KM, Levy F, Keverne EB (1992) Changes in the sensory processing of olfactory signals induced by birth in sheep. *Science,* **256**: 833–836.

Kennerly TE (1954) Local differentiation in the pocket gopher (*Geomys personatus*) in southern Texas. *Texas Journal of Science,* **6**: 297-329.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science,* **309**: 1850–1854.

Kimoto H, Haga S, Sato K, Touhara K (2005) Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons. *Nature,* **437**: 898–901.

Kimoto H, Sato K, Nodari F, Haga S, Holy TE, Touhara K (2007) Sex- and strain-specific expression and vomeronasal activity of mouse ESP family peptides. *Current Biology,* **17**: 1879–1884.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**: 111-120.

Kimura M (1983) *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge, UK.

Kimura M, Ohta T (1969) The average number of generations until fixation of a mutant gene in a finite population. *Genetics,* **61**: 763–771.

University of Cape Town

Kishida T, Kubota S, Shirayama Y, Fukami H (2007) The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biology Letters,* **3:** 428–430.

Kishida T (2008) Pattern of the divergence of olfactory receptor genes during tetrapod evolution. *PLoS ONE,* **3**: e2385.

Kishida T and Hikida T (2010) Degeneration patterns of the olfactory receptor genes in sea snakes. *Journal of evolutionary biology,* **23**: 302-310.

Kleiman DG (1974) Patterns of behavior in hystricomorph rodents. *Symposia of the Zoological Society of London***, 34**: 171-209.

Klein J (1987) Origin of major histocompatibility complex polymorphism – the transspecies hypothesis. *Human Immunology,* **19**: 155-162.

Klein J, Ono H, Klein D, O'hUigin C (1993) The accordion model of MHC evolution. *Progress in Immuno*logy, **8**:137–143.

Klein J, Horejsi V (1997) *Immunology.* Second edition, Blackwell Sci, London, UK.

Klein J, Sato A, Nagl S, O'hUigín C (1998) Molecular Trans-Species Polymorphism. *Annual Review of Ecology and Systematics,* **29**: 1-21.

Kobilka B (1992) Adrenergic receptors as models for G protein-coupled receptors. *Annal Review of Neuroscience,* **15**: 87–114.

Kock D, Ingram CM, Frabotta LJ, Honeycutt RL, Burda H (2006) On the nomenclature of Bathyergidae and *Fukomys* n. gen. (Mammalia: Rodentia). *Zootaxa,* **1142**: 51–55.

Kogan JH, Frankland PW, Silva AJ (2000) Long-term memory underlying hippocampus-dependent social recognition in mice. *Hippocampus,* **10**: 45–56.

Kondo R, Kaneko S, Sun H, Sakaizumi M, Chigusa SI (2002) Diversification of olfactory receptor genes in the Japanese medaka fish, *Oryzias latipes*. *Gene*, **282:** 113-120.

Krautwurst D, Yau KW, Reed RR (1998) Identification of ligands for olfactory receptors by functional expression of a receptor library. *Cell,* **95**: 917–926.

Krieger J, Schmitt A, Lobel D, Gudermann T, Schultz G, Breer H, Boekhoff I (1999) Selective activation of G protein subtypes in the vomeronasal organ upon stimulation with urine-derived compounds. *Journal of Biological Chemistry,* **274**: 4655–4662.

Kristiansen K (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacology & Therapeutics,* **103:** 21– 80.

Kundu S and Faulkes CG (2004) Patterns of MHC selection in African mole-rats, family Bathyergidae: the effects of sociality and habitat. *Proceedings of the Royal society of London B,* **271:** 273–278.

Lacey EA and Sherman PW (2005) Redifining eusociality: concepts, goals and levels of analysis. *Annales Zoologici Fennici*, **42**: 573-577.

University of Cape Town

Lai PC, Bahl G, Gremigni M, Matarazzo V, Clot-Faybesse O, Ronin C, Crasto CJ (2008) An olfactory receptor pseudogene whose function emerged in humans: a case study in the evolution of structure–function in GPCRs. *Journal of Structural and Functional Genomics,* **9**:29–40.

Landry C, Garant D, Duchesne P, Bernatchez L (2001) 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proceedings of the Royal society of London B,* **268**: 1279–1285.

Lane RP, Cutforth T, Axel R, Hood L, Trask BJ (2002) Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proceedings of the National Academy of Sciences of the United States of America,* **99:** 291–296.

Lange S, Neumann B, Hagemeyer P, Burda H (2005). Kairomone-guided food location in subterranean Zambian mole-rats (Cryptomys spp., Bathyergidae). *Folia Zoologica,* **54**: 263–268.

Langley WM (1988) Spiny mouse (*Acomys Cahirinus*) use of its distance senses in prey localization. *Behavioural Processes,* **16**: 67-73.

Lapidot M, Pilpel Y, Gilad Y, Falcovitz A, Sharon D, Haaf T, Lancet D (2001) Mouse-human orthology relationships in an olfactory receptor gene cluster. *Genomics,* **71**: 296–306.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007). Clustal W and Clustal X version 2.0. *Bioinformatics Applications Note,* **23:** 2947–2948.

Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB (2004) Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. *Neuron,* **43:** 703–714.

Lavocat R (1974) What is an hystricomorph? *Symposia of the Zoological Society of London,* **34**: 7-20.

Leinders-Zufall T, Lane AP, Puche AC, Ma W, Novotny MV, Shipley MT, Zufall F (2000) Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature,* **405:** 792–796.

Leinders-Zufall T, Brennan P, Widmayer P, Chandramani SP, Maul-Pavicic A, Jäger M, Li X-H, Breer H, Zufall F, Boehm T (2004) MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science,* **306**: 1033–1037.

Lessa EP, Vassallo AI, Verzi DH, Mora MS (2008) Evolution of morphological adaptations for digging in living and extinct ctenomyid and octodontid rodents. *Biological Journal of the Linnean Society*, **95**: 267–283.

Levai O, Feistel T, Breer H, Strotmann J (2006) Cells in the vomeronasal organ express odorant receptors but project to the accessory olfactory bulb. *Journal of Comparative Neurology,* **498**: 476–490.

Levring T and Fish GR (1956) The penetration of light in some tropical East African waters. *Oikos,* **7:** 98–109.

University of Cape Town

Levy F, Locatelli A, Piketty V, Tillet Y, Poindron P (1995) Involvement of the main but not the accessory olfactory system in maternal-behavior of primiparous and multiparous ewes. *Physiology & Behavior*, **57**: 97–104.

Leypold B, Yu C, Leinders-Zufall T, Kim M, Zufall F, Axel R (2002) Altered sexual and social behaviors in trp2 mutant mice. *Proceedings of the National Academy of Sciences of the United States of America,* **99**: 6376–6381.

Li WH, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. *Nature*, **292**: 237–239.

Li WH (1997) *Molecular Evolution*. Sinauer, Sunderland, MA.

Li WH, Gojobori T and Nei M (1981) Pseudogenes as a paradigm of neutral evolution. *Nature,* **292:** 237–239.

Li X, Li W, Wang H, Cao J, Maehashi K, Huang L, Bachmanov AA, Reed DR, Legrand-Defretin V, Beauchamp GK, *et al.* (2005). Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. *PLoS Genetics*, **1**: 27–35.

Liberles SD and Buck LB (2006) A second class of chemosensory receptors in the olfactory epithelium. *Nature,* **442:** 645–650.

Librado P and Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* **25**: 1451-1452.

Lim MM, Wang Z, Olazabal DE, Ren X, Terwilliger EF, Young LJ (2004) Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature,* **429**: 754–757.

Lindemann L, Ebeling M, Kratochwil NA, Bunzow JR, Grandy DK, Hoener MC (2005) Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. *Genomics,* **85**: 372–385.

Loconto J, Papes F, Chang E, Stowers L, Jones EP, Takada T, Kumanovics A, Fischer LK, Dulac C (2003) Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell,* **112:** 607–618.

Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R (2006) Interchromosomal interactions and olfactory receptor choice. *Cell,* **126**: 403–413.

Lovegrove BG and Wissel C (1988) Sociality in molerats - metabolic scaling and the role of risk sensitivity. *Oecologia,* **74**: 600-606.

Lundstrom JN and Olsson MJ (2005). Subthreshold amounts of social odorant affect mood, but not behavior, in heterosexual women when tested by a male, but not a female, experimenter. *Biological Psychology,* **70**: 197–204.

Ma M (2007) Encoding olfactory signals via multiple chemosensory systems. *Critical Reviews in Biochemistry and Molecular Biology,* **42**: 463–480.

Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial receptor codes for odors. *Cell,* **96**: 713–723.

University of Cape Town

Malnic B, Godfrey PA, Buck L (2004) The Human Olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America,* **101**: 2584-2589.

Malnic B (2007) Searching for the ligands of odorant receptors. *Molecular Neurobiology,* **35:** 175–181.

Man O, Gilad Y and Lancet D (2004) Prediction of the odorant binding site of olfactory receptor proteins by human–mouse comparisons. *Protein Science*, **13**: 240-254.

Markow TA (1996) Evolution of *Drosophila* mating systems. *Evolutionary Biology*, **29**: 73–106.

Markow TA and O'Grady PM (2005) Evolutionary genetics of reproductive behavior in *Drosophila*: connecting the dots. *Annual Review of Genetics*, **39**: 263–291.

Martini S, Silvotti L, Shirazi A, Ryba NJ, Tirindelli R (2001) Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ. *Journal of Neuroscience,* **21:** 843–848.

Mathis A and Vincent F (2000) Differential use of chemical cues in predator recognition and threat-sensitive predator-avoidance responses by larval newts (*Notophtalamus viridescens*). *Canadian Journal of Zoology,* **78**: 1646-1652.

Matochick JA (1998) Role of the main olfactory system in recognition between individual spiny mice. *Physiology & Behavior*, **42**: 217–222.

Matsui A, Go Y, Niimura Y (2010) Degeneration of Olfactory Receptor gene repertories in primates: no direct link to full trichromatic vision. *Molecular Biology and Evolution*, **27**: 1192–1200.

Matsunami H and Buck LB (1997) A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell,* **90**: 775–784.

Mayer WE, Ohuigin C, Zaleskarutczynska Z, Klein J (1992) Transspecies Origin of MHC-*DRB* polymorphism in the chimpanzee. *Immunogenetics,* **37**: 12-23.

Mayrose I and Otto SP (2011) A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. *Molecular Biology and Evolution*, **28**: 759-770.

McAuliffe K and Whitehead H (2005) Eusociality, menopause and information in matrilineal whales. *Trends in Ecology and Evolution*, **20**: 650.

McGowen MR, Clark C, Gatesy J (2008) The vestigial olfactory receptor subgenome of odontocete whales: phylogenetic congruence between gene-tree reconciliation and supermatrix methods. *Systematic Biology,* **57**: 574–590.

Menashe I, Man O, Lancet D, Gilad Y (2003) Different noses for different people. *Nature Genetics*, **34:**143-144.

Meyerhans A, Vartanian J-P, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Research*, **18**: 1687-1691.

Michener CD (1969) Comparative social behaviour of bees. *Annual Review of Entomology*, **14**: 277-342

University of Cape Town

Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology Evolution and Systematics*, **37:** 159–186.

Miyamichi K, Serizawa S, Kimura HM, Sakano H (2005) Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *Journal of Neuroscience,* **25:** 3586–3592.

Mombaerts P (1999a) Molecular Biology of Odorant Receptors in Vertebrates. *Annual Review of Neuroscience,* **22**: 487-509.

Mombaerts P (1999b) Seven-transmembrane proteins as odorant and chemosensory receptors. *Science,* **286**: 707-711.

Mombaerts P (2004) Genes and ligands for odorant, vomeronasal and taste receptors. *Nature Reviews Neuroscience*, **5**: 263-278.

Monclus R, Arroyo M, Valencia A, Miguel FJ (2009) Red foxes (*Vulpes vulpes*) use rabbit (*Oryctolagus cuniculus*) scent marks as territorial marking sites. *Journel of Ethology,* **27**: 153-156.

Müller-Schwarze D (2006) *Chemical Ecology of Vertebrates*. Cambridge University Press, Cambridge, UK.

Munday PL, Dixson DL, Donelson JM, Jones GP, Pratchett MS, Devitsina GV, Døving KB (2009) Ocean acidification impairs olfactory discrimination and homing ability of a marine fish. *Proceedings of the National Academy of Sciences of the United States of America,* **106**:1848–1852.

Munger SD, Leinders-Zufall T, Zufall F (2009) Subsystem organization of the mammalian sense of smell. *Annual Review of Physiology,* **71:** 115–140.

Mykytowy R and Goodrich BS (1974) Skin glands as organs of communication in mammals. *Journal of Investigative Dermatology*, **62**: 124–31.

Natynczuk SE and Macdonald DW (1994) Scent, sex, and the self-calibrating rat. *Journal of Chemical Ecology*, **20**: 1843–1857.

Nedbal MA, Allard MW, Honeycutt RL (1994) Molecular systematics of hystricognath rodents: evidence from the mitochondrial 12S rRNA gene. *Molecular Phylogenetics and Evolution,* **3**: 206-220.

Neff BD and Pitcher TE (2005) Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology*, **14:** 19–38.

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, **94**: 7799–7806.

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, **39**: 121–152.

Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Reviews Genetics,* **9**: 951-963.

University of Cape Town

Nevo E (1979) Adaptive convergence and divergence of subterranean mammals. *Annual Review of Ecology and Systematics,* **10**: 269-308.

Ngai J, Dowling MM, Buck L, Axel R, Chess A (1993) The family of genes encoding odorant receptors in the channel catfish. *Cell,* **72**: 657-666.

Nielsen R and Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**: 929-936.

Niimura Y and Nei M (2003) Evolution of olfactory receptor genes in the human genome. *Proceedings of the National Academy of Sciences of the United States of America,* **100**: 12235-12240.

Niimura Y and Nei M (2005a) Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene,* **346**: 13-21.

Niimura Y and Nei M (2005b) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences of the United States of America,* **102**: 6039-6044.

Niimura Y and Nei M (2006) Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *Journal of Human Genetics,* **51**: 505–517.

Nimura Y and Nei M (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE,* **8:** e708.

Niimura Y (2009) On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biology and Evolution,* **1**: 34–44.

Nishizumi H, Kumasaka K, Inoue N, Nakashima A, Sakano H (2007) Deletion of the core-H region in mice abolishes the expression of three proximal odorant receptor genes in *cis*. *Proceedings of the National Academy of Sciences of the United States of America,* **104**: 20067–20072.

Norlin E, Gussing F, Berghard A (2003) Vomeronasal phenotype and behavioral alterations in G alpha i2 mutant mice. *Current Biology,* **13**: 1214–1219.

O'Connor TD and Mundy NI (2009) Genotype-phenotype associations: substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics*, **25**: i94-i100.

Olsson M, Madsen T, Nordby J, Wapstra E, Ujvari B, Wittsell H (2003) Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal society of London B,* **270**: 254–256.

Opazo JC (2005) A molecular timescale for caviomorph rodents (Mammalia, Hystricognathi). *Molecular Phylogenetics and Evolution*, **37**: 932-937.

Opazo JC, Palma RE, Melo F, Lessa EP (2008) Adaptive evolution of the insulin gene in caviomorph rodents. *Molecular Biology and Evolution*, **22**: 1290-1298.

O'Riain, M.J. and Jarvis J.U.M (1997) Colony member recognition and xenophobia in the naked mole rat. *Animal Behaviour,* **53:** 487-498.

University of Cape Town

Orr JC, Fabry VJ, Aumont O, Bopp L, Doney SC, Feely RA, Gnanadesikan A, Gruber N, Ishida A, Joos F, *et al.* (2010) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature,* **437**: 681–686.

Otaki JM, Yamamoto H, Firestein S (2003) Odorant receptor expression in the mouse cerebral cortex. *Journal of Neurobiology,* **58:** 315–327.

Pace U, Hanski E, Salomon Y, Lancet D (1985) Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature,* **316**: 255–258.

Parada A, D'Elia G, Bidau CJ, Lessa EP (2011) Species groups and the evolutionary diversification of tuco-tucos, genus Ctenomys (Rodentia: Ctenomyidae). *Journal of Mammalogy*, **92**: 671–682.

Park TJ, Comer C, Carol A, Lu Y, Hong H-S, Rice FL (2003) Somatosensory organization and behavior in naked mole-rats: II. Peripheral structures, innervation, and selective lack of neuropeptides associated with thermoregulation and pain. *Journal of Comparative Neurology,* **465**: 104–120.

Parmentier M, Libert F, Schurmans S, Schiffmann S, Lefort A, Eggerickx D, Ledent C, Mollereau C, Gerard C, Perret J, *et al.* (1992) Expression of members of the putative olfactory receptor gene family in mammalian germ cells. *Nature,* **355**: 453–455.

Parsegian VA (2006) *van der Waals Forces: a Handbook for Biologists, Chemists, Engineers, and Physicists*. Cambridge University Press, Cambridge, UK.

Penn D and Potts WK (1998) Untrained mice discriminate MHC-determined odors. *Physiology & Behavior,* **64**: 235–243.

Penn DJ and Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist,* **153**: 145–164.

Pérez-Bercoff Å, Koch J, Bürglin TR (2006). LogoBar: bar graph visualization of protein logos with gaps. *Bioinformatics,* **22**: 112-114.

Pepper JW, Braude SH, Lacey EA, Sherman PW (1991) Vocalizations of the naked mole-rat. In *The biology of the Naked Mole-Rat.* Sherman PW, Jarvis JUM, Alexander RD (eds), Princeton University Press, Princeton, NJ, pp 243-274.

Pettigrew JD (1999) Electroreception in monotremes. *Journal of Experimental Biology,* **202:** 1447–1454.

Pin JP, Galvez T, Prezeau L (2003) Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacology and Therapeutics,* **98:** 325–354.

Popik P, Vetulani J, Bisaga A, van Ree JM (1991) Recognition cue in the rat's social memory paradigm. *Journal of Basic and Clinical Physiology and Pharmacology*, **2**: 315–327.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, **25**: 1253-1256.

Potts WK, Manning CJ, Wakeland EK (1991) Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature,* **352**: 619–621.

University of Cape Town

Pyrski M, Xu Z, Walters E, Gilbert DJ, Jenkins NA, Copeland NG, Margolis FL (2001) The OMP-lacZ transgene mimics the unusual expression pattern of OR-Z6, a new odorant receptor gene on mouse chromosome 6: implication for locus-dependent gene expression. *Journal of Neuroscience,* **21:** 4637–4648.

Qasba P and Reed RR (1998) Tissue and zonal-specific expression of an olfactory receptor transgene. *Journal of Neuroscience,* **18:** 227–236.

Quignon P, Giraud M, Rimbault M, Lavigne P, Tacher S, Morin E, Retout E, Valin A, Lindblad-Toh K, Nicolas J, Galibert F (2005) The dog and rat olfactory receptor repertoires. *Genome Biology*, **6**: R83.

R Development Core Team (2008) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0*, URL http://www.R-project.org.

Raguso RA and Willis MA (2002) Synergy between visual and olfactory cues in nectar feeding by naïve hawkmoths. *Animal Behaviour,* **63**: 685–695.

Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD (2008) Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Molecular Biology and Evolution,* **25**: 207–219.

Reeve HK and Sherman PW (1991) Intracolonial aggression and nepotism by the breeding female naked mole-rat. In *The Biology of the Naked Mole-Rat.* Sherman PW, Jarvis JUM, Alexander RD (eds), Princeton University Press, Princeton, NJ, pp 337-357.

Reeve HK, Sherman PW, Keller L (1996) The eusociality continuum revisited. *Trends in Ecology and Evolution*, **11**: 472.

Ressler KJ, Sullivan SL, Buck LB (1993) A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell,* **73**: 597–609.

Reusch TBH, Haberli MA, Aeschlimann PB, Milinski M (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature*, **414:** 300–302.

Richardson DS and Westerdahl H (2003) MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbles. *Molecular Ecology,* **12**: 3523-3529.

Riviere S, Challet L, Fluegge D, Spehr M, Rodriguez I (2009) Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature*, **459**: 574–577.

Roberts A (1951) *The Mammals of South Africa*. Trustees of the 'Mammals of South Africa' book fund, Central News Agency, Cape Town, South Africa, p 700.

Rodriguez I, Feinstein P, Mombaerts P (1999) Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell,* **97**: 199–208.

Rodriguez I, Greer CA, Mok MY, Mombaerts P (2000) A putative pheromone receptor gene expressed in human olfactory mucosa. *Nature Genetics,* **26**: 18–19.

University of Cape Town

Rodriguez I, Del Punta K, Rothman A, Ishii T, Mombaerts P (2002) Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nature Neuroscience,* **5:** 134–140.

Rohland N, Malaspinas A-S, Pollack JL, Slatkin M, Matheus P, Hofreiter M (2007) Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biology*, **5**: e207.

Rosen GJ, De Vries GJ, Goldman SL, Goldman BD, Forger NG (2007) Distribution of vasopressin in the brain of the eusocial naked mole-rat. *Journal of Comparative Neurology,* **500**: 1093–1105.

Rouquier S, Taviaux S, Trask BJ, Brand-Arpon V, van den Engh G, Demaille J, Giorgi D (1998) Distribution of olfactory receptor genes in the human genome. *Nature Genetics*, **18:** 243-250.

Rouquier S, Blancher A, Giorgi D (2000) The olfactory receptor gene repertoire in primates and mouse: Evidence for reduction of the functional fraction in primates. *Proceedings of the National Academy of Sciences of the United States of America*, **97:** 2870-2874.

Ryba NJ and Tirindelli R (1997) A new multigene family of putative pheromone receptors. *Neuron,* **19**: 371–379.

Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD (2009) Odor coding by a mammalian receptor repertoire. *Science Signaling,* **2:** ra9.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**: 406–425.

Salazar I, Barber PC, Cifuentes JM (1992) Anatomical and immunohistological demonstration of the primary neural connections of the vomeronasal organ in the dog. *The Anatomical Record,* 233: **309–313.**

Sam M, Vora S, Malnic B, Ma W, Novotny MV, Buck LB (2001) Neuropharmacology. Odorants may arouse instinctive behaviour. *Nature,* **412:** 142.

Sambrook J, Frisch E, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, volumes **1-3**.

Sanchez-Andrade G and Kendrick KM (2009) The main olfactory system and social learning in mammals. *Behavioural Brain Research,* **200:** 323–335.

Sanders KL, Lee MSY, Leys R, Foster R, Keogh JS (2008) Molecular phylogeny and divergence dates for Austral-asian elapids and sea snakes (hydrophiinae): evidence from seven genes for rapid evolutionary radiations. *Journal of Evolutionary Biology,* **21**: 682–695.

Sarafoleanu C, Mella C, Georgescu M, Perederco C (2009) The importance of the olfactory sense in the human behavior and evolution. *Journal of Medicine and Life* **2**: 196-198.

Savic I, Berglund H, Gulyas B, Roland P (2001) Smelling of odorous sex hormone-like compounds causes sex-differentiated hypothalamic activations in humans. *Neuron,* **31**: 661-668.

Schilling A (1979) Olfactory communication in prosimians. *In: Doyle GA, Martin RD (eds) The study of prosimian behavior. Academic Press, New York,* 461-542.

University of Cape Town

Schluter D and Price T (1993) Honesty, perception and population divergence in sexually selected traits. *Proceedings of the Royal society of London B*, **253:** 117–122.

Seehausen O, van Alphen JJM, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science,* **277**: 1808–1811.

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HD, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature,* **455**: 620–626.

Serizawa S, Miyamichi K, Nakatani H, Suzuki M, Saito M, Yoshihara Y, Sakano H (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science,* **302**: 2088–2094.

Sharon D, Glusman G, Pilpel Y, Khen M, Gruetzner F, Haaf T, Lancet D (1999) Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics*, **61:** 24-36.

Sherman PW, Lacey EA, Reeve HK, Keller L (1995) The eusociality continuum. *Behavioral Ecology*, **6**: 102-108.

Shi P, Bielawski JP, Yang H, Zhang YP (2005) Adaptive diversification of vomeronasal receptor 1 genes in rodents. *Journal of Molecular Evolution,* **60**: 566–576.

Shirokova E, Raguse JD, Meyerhof W, Krautwurst D (2008) The human vomeronasal type-1 receptor family—detection of volatiles and cAMP signaling in HeLa/Olf cells. *FASEB Journal,* **22**: 1416–1425.

Silvotti L, Giannini G, Tirindelli R (2005) The vomeronasal receptor V2R2 does not require escort molecules for expression in heterologous systems. *Chemical Senses,* **30:** 1–8.

Silvotti L, Moiani A, Gatti R, Tirindelli R (2007) Combinatorial co-expression of pheromone receptors, V2Rs. *Journal of Neurochemistry,* **103:** 1753–1763.

Simmons NB, Seymour KL, Habersetzer J, Gunnell GF (2008) Primitive early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature,* **451**: 818–821.

Singer MS, Weisinger-Lewin Y, Lancet D, Shepherd GM (1996) Positive selection moments identify potential functional residues in human olfactory receptors. *Receptors Channels,* **4:** 141–147.

Sklar PB, Anholt RRH, Snyder SH (1986) The odorant-sensitive adenylate cyclase of olfactory receptor cells: differential stimulation by distinct classes of odorants. *Journal of Biological Chemistry,* **261**: 15538–15543.

Smith TD, Bhatnagar KP, Dennis JC, Morrison EE, Park TJ (2007) Growth deficient vomeronasal organs in the naked mole-rat (*Heterocephalus glaber*). *Brain Research*, **1132**: 78-83.

Sosinsky A, Glusman G, Lancet D (2000) The genomic structure of human olfactory receptor genes. *Genomics,* **70**: 49–61.

University of Cape Town

Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzel CH, Zimmer RK, Hatt H (2003) Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science,* **299**: 2054–2058.

Spehr M, Schwane K, Riffell JA, Zimmer RK, Hatt H. (2006a) Odorant receptors and olfactory-like signaling mechanisms in mammalian sperm. *Molecular and Cellular Endocrinology,* **250:** 128–136.

Spehr M, Spehr J, Ukhanov K, Kelliher KR, Leinders-Zufall T, Zufall F (2006b) Parallel processing of social signals by the mammalian main and accessory olfactory systems. *Cellular and Molecular Life Sciences,* **63:** 1476–1484.

Spinks AC, Jarvis JUM, Bennett NC (2000) Comparative patterns of philopatry and dispersal in two common mole-rat populations: implications for the evolution of mole-rat sociality. *Journal of Animal Ecology,* **69**: 224-234.

Steiger SS, Fidler AE., Valcu M, Kempenaers B (2008) Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? *Proceedings of the Royal society of London B*, **275**: 2309-2317.

Steiger SS, Fidler AE, Kempenaers B (2009) Evidence for increased olfactory receptor gene repertoire size in two nocturnal bird species with well-developed olfactory ability. *BMC Evolutionary Biology,* **9**: 117.

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research,* **35**: 506-511.

Stopka P, Janotova K, Heyrovsky D (2007) The advertisement role of major urinary proteins in mice. Physiology & Behavior, **91**: 667–670.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, **64**: 479-498.

Storey JD (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, **31**: 2013-2035.

Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B*, **66**: 187-205.

Stowers L, Holy T, Meister M, Dulac C, Koentges G (2002) Loss of sex discrimination and male–male aggression in mice deficient for TRP2. *Science,* **295**: 1493–1500.

Strotmann J, Wanner I, Krieger J, Raming K, Breer H (1992) Expression of odorant receptors in spatially restricted subsets of chemosensory neurones. *Neuroreport,* **3**: 1053–1056.

Strotmann J, Conzelmann S, Beck A, Feinstein P, Breer H, Mombaerts P (2000) Local permutations in the glomerular array of the mouse olfactory bulb. *Journal of Neuroscience,* **20**: 6927–6938.

Sullivan SL, Adamson MC, Ressler KJ, Kozak CA, Buck LB (1996) The chromosomal distribution of mouse odorant receptor genes. *Proceedings of the National Academy of Sciences of the United States of America*, **93:** 884-888.

University of Cape Town

Swaisgood RR, Lindburg DG, Zhou X (1999) Giant pandas discriminate individual differences in conspecific scent. *Animal Behaviour,* **57**: 1054–1053.

Swofford DL (2000) *PAUP*. Phylogenetic Analysis Using Parsimony* (* and Other Methods). Sinauer Associates, Sunderland, Mass.

Tacher S, Quignon P, Rimbault M, Dreano S, Andre C, Galibert F, Quignon P, Giraud M, Rimbault M, Lavigne P, *et al.* (2005) Olfactory receptor sequence polymorphism within and between breeds of dogs. *Journal of Heredity*, **96:** 812-816.

Taga ME and Bassler BL (2003) Chemical communication among bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* **100**: 14549–14554.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* **123**: 585-595.

Takahata N and Nei M (1990) Allelic genealogy under overdominand and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics,* **124**: 967-978.

Tamura K and Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution,* **10**: 512-526.

Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**: 1596-1599.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011). *MEGA5*: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution (In Press).*

Taniguchi K and Mikami S (1985) Fine structure of the epithelia of the vomeronasal organ of horse and cattle. A comparative study. *Cell and Tissue Research,* **240**: 41- 48.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* (American Mathematical Society), **17**: 57–86.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ (2007) A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science,* **307**: 580-584.

Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HDJ, *et al.* (2006) Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biology,* **4**: 2244–2251.

Terakita A (2005) The Opsins. *Genome Biology,* **6**: 213.

The Royal Society (2005) *Ocean Acidification Due to Increasing Atmospheric Carbon Dioxide, The Royal Society, London.*

Thewissen JGM, George J, Rosa C, Kishida T (2011) Olfaction and brain size in the bowhead whale. *Marine Mammal Science*, **27**: 282–294.

University of Cape Town

Thigpen LW (1940) Histology of the skin of a normally hairless rodent. *Journal of Mammalogy,* **21**: 449-456.

Thompson JD, Higgins DG, Gibson TJ (1994) *CLUSTAL W*: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* **22**: 4673–4680.

Tirindelli R, Mucignat-Caretta C, Ryba NJ (1998) Molecular aspects of pheromonal communication via the vomeronasal organ of mammals. *Trends in Neuroscience,* **21**: 482–486.

Tirindelli R, Dibattista M, Pifferi S, Menini A (2009) From Pheromones to Behavior. *Physiological Reviews,* **89**: 921–956.

Touhara K, Sengoku S, Inaki K, Tsuboi A, Hirono J, Sato T, Sakano H, Haga T (1999) Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proceedings of the National Academy of Sciences of the United States of America,* **96:** 4040–4045.

Touhara K (2007) Molecular biology of peptide pheromone production and reception in mice. *Advances in Genetics,* **59**: 147–171.

Touhara K and Vosshall LB (2009) Sensing odorants and pheromones with chemosensory receptors. *Annual Review of Physiology,* **71**: 307 -332.

Trask BJ, Massa H, Brand-Arpon V, Chan K, Friedman C, Nguyen OT, Eichler E, van den Engh G, Rouquier S, Shizuya H, *et al.* (1998a) Large multichromosomal duplications encompass many members of the olfactory receptor gene family in the human genome. *Human Molecular Genetics*, **7:** 2007-2020.

Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, Blankenship J, Collins C, Giorgi D, Iadonato S, Johnson F, *et al.* (1998b) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Human Molecular Genetics*, **7:** 13-26.

Turner LM, Chuong EB, Hoekstra HE (2008) Comparative analysis of testis protein evolution in rodents. *Genetics*, **179**: 2075–2089.

Vaccarezza OL, Sepich LN, Tramezzani JH (1981) The vomeronasal organ of the rat. *Journal of Anatomy,* **132**: 167–185.

Van Daele APPG, Faulkes CG, Verheyen E, Adriaens D (2007) African Mole-rats (Bathyergidae): A Complex Radiation in Tropical Soils. In *Subterranean Rodents: News from Underground*. Begall S, Burda H, Schleich CE (eds), Springer-Verlag, Heidelberg, Germany, pp 357-373.

Vanderhaeghen P, Schurmans S, Vassart G, Parmentier M (1997) Molecular cloning and chromosomal mapping of olfactory receptor genes expressed in the male germ line: evidence for their wide distribution in the human genome. *Biochemical and Biophysical Research Communications,* **237**: 283–287.

Van Houten J (1994) Chemosensory transduction in eukaryotic microorganisms: trends for neuroscience*?* *Trends in Neurosciences*, **17**: 62-71.

University of Cape Town

Vassalli A, Rothman A, Feinstein P, Zapotocky M, Mombaerts P (2002) Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron,* **35:** 681–696.

Vassar R, Ngai J, Axel R (1993) Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell,* **74:** 309–318.

Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE, Gaitanaris GA (2003) The G protein-coupled receptor repertoires of human and mouse. *Proceedings of the National Academy of Sciences of the United States of America,* **100**: 4903-4908.

Volz A, Ehlers A, Younger R, Forbes S, Trowsdale J, Schnorr D, Beck S, Ziegler A (2003) Complex transcription and splicing of odorant receptor genes. *Journal of Biological Chemistry,* **278**: 19691-19701.

Wallace ED and Bennett NC (1998) The colony structure and social organization of the giant Zambian mole-rat, *Cryptomys mechowi*. *Journal of Zoology, London,* **244**: 51-61.

Wang Z, Nudelman A, Storm DR (2007) Are pheromones detected through the main olfactory epithelium? *Molecular Neurobiology,* **35**: 317–323.

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, *et al.* (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature,* **453**: 175-183.

Wcislo WT (1997) Social terminology: what are words worth? *Trends in Ecology and Evolution*, **12**: 161.

Weber M, Pehl U, Breer H, Strotmann J (2002) Olfactory receptor expressed in ganglia of the autonomic nervous system. *Journal of Neuroscience Research,* **68:** 176–184.

Whinnet A and Mundy NI (2003) Isolation of novel olfactory receptor genes in marmosets (*Callithrix*): insights into pseudogene formation and evidence for functional degeneracy in non-human primates. *Gene,* **304**: 87-96.

Whittaker RH and Feeny PP (1971) Allelochemics: chemical interactions between species. *Science,* **171**: 757-770.

Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin,* **1:** 80–83.

Wilson EO (1971) *The Insect Societies.* Harvard University Press, Cambridge, MA.

Wilson DE and Reeders DM (2005) *Mammal Species of the World. A Taxonomic and Geographic Reference*. 3rd edition, Johns Hopkins University Press (available at http://www.press.jhu.edu).

Wingfield JC (2005) A continuing saga: the role of testosterone in aggression. *Hormones and Behavior,* **48**: 253–255.

Winter EE, Goodstadt L, Ponting CP (2004) Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Research,* **14**: 54–61.

University of Cape Town

Wlasiuk G, Garza JC, Lessa EP (2003) Genetic and geographic differentiation in the Rio Negro tuco-tuco (Ctenomys rionegrensis): inferring the roles of migration and drift from multiple genetic markers. *Evolution,* **57**: 913–926.

Wolfsberg TG, Primakoff P, Myles DG, White JM (1995) ADAM, a novel family of membrane proteins containing a disintegrin and metalloprotease domain: multipotential functions in cell–cell and cell–matrix interactions. *Journal of Cell Biology* **131**: 275–278.

Wong A (2011) The molecular evolution of animal reproductive tract proteins: what have we learned from mating-system comparisons? *International Journal of Evolutionary Biology*, DOI: 10.4061/2011/908735.

Wood AE (1985) The relationships, origin, and dispersal of the hystricognathus rodents. In *Evolutionary Relationships among Rodents: A Multidisciplinary Analysis*. Luckett WP and Hartenberger J-L (eds), NATO ASI Series, New York, pp 475-513.

Wyart C, Webster WW, Chen JH, Wilson SR, McClary A, Khan RM, Sobel N (2007) Smelling a single component of male sweat alters levels of cortisol in women. *Journal of Neuroscience,* **27**: 1261–1265.

Wyatt TD (2003) *Pheromones and Animal Behavior: Communication by Smell and Taste*. Oxford University Press, Oxford, UK.

Wysocki CJ, Beauchamp GK (1984) Ability to smell androstenone is genetically determined. *Proceedings of the National Academy of Sciences of the United States of America*, **81:** 4899-4902.

Wysocki CJ and Preti G (2004) Facts, fallacies, fears, and frustrations with human pheromones. *Anatomical Record part A – Discoveries in Molecular Cellular and Evolutionary Biology,* **281**: 1201–1211.

Xie SY, Feinstein P, Mombaerts P (2000) Characterization of a cluster comprising ~100 odorant receptor genes in mouse. *Mammalian Genome,* **11:** 1070–1078.

Xu F, Schaefer M, Kida I, Schafer J, Liu N, Rothman DL, Hyder F, Restrepo D, Shepherd GM (2005) Simultaneous activation of mouse main and accessory olfactory bulbs by odors or pheromones. *Journal of Comparative Neurology*, **489**: 491–500.

Yang H, Shi P, Zhang YP, Zhang J (2005a) Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics,* **86:** 306–315.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in BioSciences,* **13**: 555-556.

Yang Z and Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution,* **19**: 908–917.

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution,* **22**: 1107-1118.

University of Cape Town

Yang Z (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586-1591.

Yang Z (2009) *PAML manual.* Version 4.3, September 2009.

Yokoyama S and Yokoyama R (1996) Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annual Review of Ecology and Systematics*, **27**: 543–567.

Yokoyama S and Shi Y (2000) Genetics and evolution of ultraviolet vision in vertebrates. *FEBS Letters*, **486**: 167–172.

Yokoyama S (2000) Molecular evolution of vertebrate visual pigments. *Progress in Retinal and Eye Research,* **19**: 385–419.

Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Molecular Genetics*, **11:** 535-546.

Young JM and Trask BJ (2002) The sense of smell: genomics of vertebrate odorant receptors. *Human Molecular Genetics,* **11**: 1153–1160.

Young JM, Shykind BM, Lane RP, Tonnes-Priddy L, Ross JA, Walker M, Williams EM, Trask BJ (2003) Odorant receptor expressed sequence tags demonstrate olfactory expression of over 4000 genes, extensive alternate splicing and unequal expression levels. *Genome Biology,* **4**: R71.

Young JM, Kambere M, Trask BJ, Lane RP (2005) Divergent V1R repertoires in five species: amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs. *Genome Research,* **15**: 231–240.

Young JM and Trask BJ (2007) V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *Trends in Genetics,* **23:** 212–215.

Zarzo M (2007) The sense of smell: Molecular basis of odorant recognition. *Biological Reviews of the Cambridge Philosophical Society,* **82**: 455-479.

Zhai W, Nielsen R, Goldman N, Yang Z (2012) Looking for Darwin in Genomic Sequences—Validity and Success of Statistical Methods. *Molecular Biology and Evolution*, doi:10.1093/molbev/mss104.

Zhang X and Firesetin S (2002) The olfactory receptor gene superfamily of the mouse. *Nature neuroscience,* **5**: 124-133.

Zhang X, Rodriguez I, Mombaerts P, Firestein S (2004) Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics,* **83**: 802-811.

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution,* **22**: 2472-2479.

Zhang X, Zhang X, Firestein S (2007a) Comparative genomics of odorant and pheromone receptor genes in rodents. *Genomics*, **89**: 441-450.

University of Cape Town

Zhang X, De la Cruz O, Pinto JM, Nicolae D, Firestein S, Gilad Y (2007b) Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biology,* **8**: R86.

Zhao H, Ivic L, Otaki JM, Hashimoto M, Mikoshiba K, Firestein S (1998) Functional expression of a mammalian odorant receptor. *Science,* **279**: 237–242.

Zhao H, Rossiter SJ, Teeling EC, Li C, Cotton JA, Zhang S (2009a) The evolution of color vision in nocturnal mammals. *Proceedings of the National Academy of Sciences of the United States of America,* **106**: 8980–8985.

Zheng D and Gerstein MB (2007) The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends in Genetics,* **23**: 219–224.

Zozulya S, Echeverri F, Nguyen T (2001) The human olfactory receptor repertoire. *Genome Biology* **2**: 0018.1-0018.12.

University of Cape Town

Appendix I Lists and Tables

I.1 Single letter amino-acid notation (*IUPAC* codes, 1971)

Single letter code	Amino-acid
A	Alanine
R	Arginine
N	Asparagine
D	Aspartic acid
C	Cysteine
Q	Glutamine
E	Glutamic acid
G	Glycine
H	Histidine
I	Isoleucine
L	Leucine
K	Lysine
M	Methionine
F	Phenylalanine
P	Proline
S	Serine
T	Threonine
W	Tryptophan
Y	Tyrosine
V	Valine

University of Cape Town

I.2 List of Bathyergid OR sequence names Abbreviations correspond to gene names used throughout this thesis for bathyergid OR genes; full names have been assigned following the explanatory nomenclature recently used by Hayden et al. (2010) in the most comprehensive study of mammalian OR genes.

List of Sequence names	
Abbreviation	**Full name**
BJ4_A1	Bathyergus janetta clone BJ4_A1 olfactory receptor family 7 (OR7) gene, partial cds
BJ4_A3_P	Bathyergus janetta clone BJ4_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BJ4_A4_P	Bathyergus janetta clone BJ4_A4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BJ4_A5_P	Bathyergus janetta clone BJ4_A5 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BJ4_A6_P	Bathyergus janetta clone BJ4_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BJ4_A12	Bathyergus janetta clone BJ4_A12 olfactory receptor family 7 (OR7) gene, partial cds
BS7_A1_P	Bathyergus suillus clone BS7_A1 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BS7_A2	Bathyergus suillus clone BS7_A2 olfactory receptor family 7 (OR7) gene, partial cds
BS7_A3_P	Bathyergus suillus clone BS7_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BS7_A4	Bathyergus suillus clone BS7_A4 olfactory receptor family 7 (OR7) gene, partial cds
BS7_A5	Bathyergus suillus clone BS7_A5 olfactory receptor family 7 (OR7) gene, partial cds
BS7_A6_P	Bathyergus suillus clone BS7_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BS7_A7_P	Bathyergus suillus clone BS7_A7 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BS7_A9_P	Bathyergus suillus clone BS7_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
BS7_A10	Bathyergus suillus clone BS7_A10 olfactory receptor family 7 (OR7) gene, partial cds
CA1_A3_P	Fukomys amatus clone CA1_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CA1_A5	Fukomys amatus clone CA1_A5 olfactory receptor family 7 (OR7) gene, partial cds
CA1_A8	Fukomys amatus clone CA1_A8 olfactory receptor family 7 (OR7) gene, partial cds
CA1_A12_P	Fukomys amatus clone CA1_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CA3_A8_P	Fukomys amatus clone CA3_A8 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CA3_A9	Fukomys amatus clone CA3_A9 olfactory receptor family 7 (OR7) gene, partial cds
CA3_A10	Fukomys amatus clone CA3_A10 olfactory receptor family 7 (OR7) gene, partial cds
CA3_A11	Fukomys amatus clone CA3_A11 olfactory receptor family 7 (OR7) gene, partial cds
CA3_A12	Fukomys amatus clone CA3_A12 olfactory receptor family 7 (OR7) gene, partial cds
CA3_B1_P	Fukomys amatus clone CA3_B1 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CA3_B3_P	Fukomys amatus clone CA3_B3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CA3_B4	Fukomys amatus clone CA3_B4 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A2	Fukomys anselli clone CAN3_A2 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A3	Fukomys anselli clone CAN3_A3 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A4	Fukomys anselli clone CAN3_A4 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A5	Fukomys anselli clone CAN3_A5 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A6	Fukomys anselli clone CAN3_A6 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A7	Fukomys anselli clone CAN3_A7 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A9_P	Fukomys anselli clone CAN3_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CAN3_A10	Fukomys anselli clone CAN3_A10 olfactory receptor family 7 (OR7) gene, partial cds
CAN3_A11	Fukomys anselli clone CAN3_A11 olfactory receptor family 7 (OR7) gene, partial cds
CAN4_A11_P	Fukomys anselli clone CAN3_A11 olfactory receptor family 7 (OR7) pseudogene, partial sequence

CAN4_A12_P	Fukomys anselli clone CAN3_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CAN4_B4	Fukomys anselli clone CAN4_B4 olfactory receptor family 7 (OR7) gene, partial cds
CAN4_B5_P	Fukomys anselli clone CAN4_B5 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CAN4_B6_P	Fukomys anselli clone CAN4_B6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CAN4_B7_P	Fukomys anselli clone CAN4_B7 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CAN4_B9	Fukomys anselli clone CAN4_B9 olfactory receptor family 7 (OR7) gene, partial cds
CAN4_B10_P	Fukomys anselli clone CAN4_B10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CAN4_B11	Fukomys anselli clone CAN4_B11 olfactory receptor family 7 (OR7) gene, partial cds
CAN4_B12_P	Fukomys anselli clone CAN4_B12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB1_A2_P	Fukomys bocagei clone CB1_A2 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB1_A3_P	Fukomys bocagei clone CB1_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB1_A4_P	Fukomys bocagei clone CB1_A4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB1_A10_P	Fukomys bocagei clone CB1_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB1_A12	Fukomys bocagei clone CB1_A12 olfactory receptor family 7 (OR7) gene, partial cds
CB2_A1_P	Fukomys bocagei clone CB2_A1 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB2_A3_P	Fukomys bocagei clone CB2_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB2_A4	Fukomys bocagei clone CB2_A4 olfactory receptor family 7 (OR7) gene, partial cds
CB2_A5_P	Fukomys bocagei clone CB2_A5 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB2_A6	Fukomys bocagei clone CB2_A6 olfactory receptor family 7 (OR7) gene, partial cds
CB2_A7	Fukomys bocagei clone CB2_A7 olfactory receptor family 7 (OR7) gene, partial cds
CB2_A8_P	Fukomys bocagei clone CB2_A8 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB2_A9_P	Fukomys bocagei clone CB2_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB2_A10_P	Fukomys bocagei clone CB2_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CB2_A12	Fukomys bocagei clone CB2_A12 olfactory receptor family 7 (OR7) gene, partial cds
CD7_A2_P	Fukomys darlingi clone CD7_A2 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7_A6	Fukomys darlingi clone CD7_A6 olfactory receptor family 7 (OR7) gene, partial cds
CD7_A7_P	Fukomys darlingi clone CD7_A7 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7_A8_P	Fukomys darlingi clone CD7_A8 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_A3_P	Fukomys darlingi clone CD7b_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_A6_P	Fukomys darlingi clone CD7b_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_A9_P	Fukomys darlingi clone CD7b_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_A11_P	Fukomys darlingi clone CD7b_A11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_A12_P	Fukomys darlingi clone CD7b_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_B3_P	Fukomys darlingi clone CD7b_B3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CD7b_B5_P	Fukomys darlingi clone CD7b_B5 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM4_A2	Fukomys damarensis clone CDM4_A2 olfactory receptor family 7 (OR7) gene, partial cds
CDM4_A3_P	Fukomys damarensis clone CDM4_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM4_A4_P	Fukomys damarensis clone CDM4_A4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM4_A6_P	Fukomys damarensis clone CDM4_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM4_A9_P	Fukomys damarensis clone CDM4_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM6_A4	Fukomys damarensis clone CDM6_A4 olfactory receptor family 7 (OR7) gene, partial cds
CDM6_A5	Fukomys damarensis clone CDM6_A5 olfactory receptor family 7 (OR7) gene, partial cds
CDM6_A7	Fukomys damarensis clone CDM6_A7 olfactory receptor family 7 (OR7) gene, partial cds
CDM6_A8	Fukomys damarensis clone CDM6_A8 olfactory receptor family 7 (OR7) gene, partial cds
CDM6_C5	Fukomys damarensis clone CDM6_C5 olfactory receptor family 7 (OR7) gene, partial cds

CDM6_C10	Fukomys damarensis clone CDM6_C10 olfactory receptor family 7 (OR7) gene, partial cds
CDM8_A3	Fukomys damarensis clone CDM8_A3 olfactory receptor family 7 (OR7) gene, partial cds
CDM8_A6	Fukomys damarensis clone CDM8_A6 olfactory receptor family 7 (OR7) gene, partial cds
CDM8_A9_P	Fukomys damarensis clone CDM8_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM8_A10_P	Fukomys damarensis clone CDM8_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CDM8_A12_P	Cryptomys damarensis clone CDM8_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHH10_A1	Cryptomys hottentotus hottentotus clone CHH10_A1 olfactory receptor family 7 (OR7) gene, partial cds
CHH10_A2	Cryptomys hottentotus hottentotus clone CHH10_A2 olfactory receptor family 7 (OR7) gene, partial cds
CHH10_A3_P	Cryptomys hottentotus hottentotus clone CHH10_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHH10_A4	Cryptomys hottentotus hottentotus clone CHH10_A4 olfactory receptor family 7 (OR7) gene, partial cds
CHH10_A6	Cryptomys hottentotus hottentotus clone CHH10_A6 olfactory receptor family 7 (OR7) gene, partial cds
CHH10_A7	Cryptomys hottentotus hottentotus clone CHH10_A7 olfactory receptor family 7 (OR7) gene, partial cds
CHH10_A8_P	Cryptomys hottentotus hottentotus clone CHH10_A8 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHH10_A9_P	Cryptomys hottentotus hottentotus clone CHH10_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHH10_A10	Cryptomys hottentotus hottentotus clone CHH10_A10 olfactory receptor family 7 (OR7) gene, partial cds
CHH10_A11_P	Cryptomys hottentotus hottentotus clone CHH10_A11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHH10_A12	Cryptomys hottentotus hottentotus clone CHH10_A12 olfactory receptor family 7 (OR7) gene, partial cds
CHN4_A1_P	Cryptomys hottentotus natalensis clone CHN4_A1 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHN4_A2_P	Cryptomys hottentotus natalensis clone CHN4_A2 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHN4_A4_P	Cryptomys hottentotus natalensis clone CHN4_A4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHN4_A5	Cryptomys hottentotus natalensis clone CHN4_A5 olfactory receptor family 7 (OR7) gene, partial cds
CHN4_A7	Cryptomys hottentotus natalensis clone CHN4_A7 olfactory receptor family 7 (OR7) gene, partial cds
CHN4_A8	Cryptomys hottentotus natalensis clone CHN4_A8 olfactory receptor family 7 (OR7) gene, partial cds
CHN4_A10_P	Cryptomys hottentotus natalensis clone CHN4_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHN4_A11	Cryptomys hottentotus natalensis clone CHN4_A11 olfactory receptor family 7 (OR7) gene, partial cds
CHP2_A5	Cryptomys hottentotus pretoriae clone CHP2_A5 olfactory receptor family 7 (OR7) gene, partial cds
CHP2_A6	Cryptomys hottentotus pretoriae clone CHP2_A6 olfactory receptor family 7 (OR7) gene, partial cds
CHP2_A9	Cryptomys hottentotus pretoriae clone CHP2_A9 olfactory receptor family 7 (OR7) gene, partial cds
CHP2_A10_P	Cryptomys hottentotus pretoriae clone CHP2_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_A3_P	Cryptomys hottentotus pretoriae clone CHP3_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_A7_P	Cryptomys hottentotus pretoriae clone CHP3_A7 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_A10_P	Cryptomys hottentotus pretoriae clone CHP3_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_A11_P	Cryptomys hottentotus pretoriae clone CHP3_A11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_A12	Cryptomys hottentotus pretoriae clone CHP3_A12 olfactory receptor family 7 (OR7) gene, partial cds
CHP3_B5_P	Cryptomys hottentotus pretoriae clone CHP3_B5 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_B6_P	Cryptomys hottentotus pretoriae clone CHP3_B6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CHP3_B11_P	Cryptomys hottentotus pretoriae clone CHP3_B11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CM6_A2	Fukomys mechowi clone CM6_A2 olfactory receptor family 7 (OR7) gene, partial cds
CM6_A3	Fukomys mechowi clone CM6_A3 olfactory receptor family 7 (OR7) gene, partial cds
CM6_A4	Fukomys mechowi clone CM6_A4 olfactory receptor family 7 (OR7) gene, partial cds
CM6_A5_P	Fukomys mechowi clone CM6_A5 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CM6_A6_P	Fukomys mechowi clone CM6_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
CM6_A7	Fukomys mechowi clone CM6_A7 olfactory receptor family 7 (OR7) gene, partial cds
CM6_A8	Fukomys mechowi clone CM6_A8 olfactory receptor family 7 (OR7) gene, partial cds
CM6_A9_P	Fukomys mechowi clone CM6_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence

CM6_A12	Fukomys mechowi clone CM6_A12 olfactory receptor family 7 (OR7) gene, partial cds
GC8_A3	Georychus capensis clone GC8_A3 olfactory receptor family 7(OR7) gene, partial cds
GC8_A6_P	Georychus capensis clone GC8_A6 olfactory receptor family 7(OR7) pseudogene, partial sequence
GC8_A11	Georychus capensis clone GC8_A11 olfactory receptor family 7(OR7) gene, partial cds
GC10_A2_P	Georychus capensis clone GC10_A2 olfactory receptor family 7(OR7) pseudogene, partial sequence
GC10_A4	Georychus capensis clone GC10_A4 olfactory receptor family 7(OR7) gene, partial cds
GC10_A5	Georychus capensis clone GC10_A5 olfactory receptor family 7(OR7) gene, partial cds
GC10_A8	Georychus capensis clone GC10_A8 olfactory receptor family 7(OR7) gene, partial cds
GC10_A9	Georychus capensis clone GC10_A9 olfactory receptor family 7(OR7) gene, partial cds
GC10_A10_P	Georychus capensis clone GC10_A10 olfactory receptor family 7(OR7) pseudogene, partial sequence
GC10_A11_P	Georychus capensis clone GC10_A11 olfactory receptor family 7(OR7) pseudogene, partial sequence
GC10_A12_P	Georychus capensis clone GC10_A12 olfactory receptor family 7(OR7) pseudogene, partial sequence
HA1_A3	Heliophobius argentocinereus clone HA1_A3 olfactory receptor family 7 (OR7) gene, partial cds
HA1_A4_P	Heliophobius argentocinereus clone HA1_A4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HA1_A6_P	Heliophobius argentocinereus clone HA1_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HA1_A7	Heliophobius argentocinereus clone HA1_A7 olfactory receptor family 7 (OR7) gene, partial cds
HA1_A8	Heliophobius argentocinereus clone HA1_A8 olfactory receptor family 7 (OR7) gene, partial cds
HA1_A9_P	Heliophobius argentocinereus clone HA1_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HA1_A10	Heliophobius argentocinereus clone HA1_A10 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A1	Heliophobius argentocinereus clone HA3_A1 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A2	Heliophobius argentocinereus clone HA3_A2 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A3	Heliophobius argentocinereus clone HA3_A3 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A4	Heliophobius argentocinereus clone HA3_A4 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A7	Heliophobius argentocinereus clone HA3_A7 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A9	Heliophobius argentocinereus clone HA3_A9 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A10	Heliophobius argentocinereus clone HA3_A10 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A11	Heliophobius argentocinereus clone HA3_A11 olfactory receptor family 7 (OR7) gene, partial cds
HA3_A12_P	Heliophobius argentocinereus clone HA3_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A2_P	Heterocephalus glaber clone HG2_A2 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A4_P	Heterocephalus glaber clone HG2_A4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A6_P	Heterocephalus glaber clone HG2_A6 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A7_P	Heterocephalus glaber clone HG2_A7 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A9_P	Heterocephalus glaber clone HG2_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A10_P	Heterocephalus glaber clone HG2_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A11_P	Heterocephalus glaber clone HG2_A11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_A12_P	Heterocephalus glaber clone HG2_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_B1_P	Heterocephalus glaber clone HG2_B1 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_B4_P	Heterocephalus glaber clone HG2_B4 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_B7	Heterocephalus glaber clone HG2_B7 olfactory receptor family 7 (OR7) gene, partial cds
HG2_B9_P	Heterocephalus glaber clone HG2_B9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG2_B11_P	Heterocephalus glaber clone HG2_B11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A1_P	Heterocephalus glaber clone HG4_A1 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A2_P	Heterocephalus glaber clone HG4_A2 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A3_P	Heterocephalus glaber clone HG4_A3 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A6	Heterocephalus glaber clone HG4_A6 olfactory receptor family 7 (OR7) gene, partial cds

HG4_A7_P	Heterocephalus glaber clone HG4_A7 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A8_P	Heterocephalus glaber clone HG4_A8 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A9_P	Heterocephalus glaber clone HG4_A9 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A10_P	Heterocephalus glaber clone HG4_A10 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A11_P	Heterocephalus glaber clone HG4_A11 olfactory receptor family 7 (OR7) pseudogene, partial sequence
HG4_A12_P	Heterocephalus glaber clone HG4_A12 olfactory receptor family 7 (OR7) pseudogene, partial sequence

University of Cape Town

I.3 List of unique OR genes Unique OR genes are listed for each bathyergid species; abbreviations correspond to gene names as per Appendix I.2.

Unique OR genes		
B. janetta	*F. darlingi*	*G. capensis*
BJ4_A12	CD7_A8_P	GC8_A11
BJ4_A1	CD7_A2_P	GC8_A3
BJ4_A3_P	CD7_A7_P	GC8_A6_P
BJ4_A4_P	CD7b_B5_P	GC10_A12_P
BJ4_A5_P	CD7b_A6_P	GC10_A2_P
B. suillus	CD7b_A11_P	GC10_A4
BS7_A1_P	CD7b_A12_P	GC10_A5
BS7_A3_P	CD7b_B3_P	GC10_A9
BS7_A4	*F. damarensis*	GC10_A10_P
BS7_A5	CDM4_A3_P	*H. argentocinereus*
BS7_A6_P	CDM4_A4_P	HA1_A10
BS7_A7_P	CDM6_A5	HA1_A3
BS7_A10	CDM6_A7	HA1_A4_P
F. amatus	CDM6_A8	HA1_A6_P
CA1_A12_P	CDM8_A12_P	HA1_A7
CA1_A3_P	CDM8_A3	HA1_A8
CA1_A8	CDM8_A9_P	HA1_A9_P
CA3_B3_P	*C. h. hottentotus*	HA3_A3
CA3_A8_P	CHH10_A12	HA3_A4
CA3_A9	CHH10_A1	HA3_A9
CA3_A11	CHH10_A2	*H. glaber*
CA3_B4	CHH10_A3_P	HG2_B11_P
F. anselli	CHH10_A4	HG2_A2_P
CAN3_A2	CHH10_A7	HG2_A4_P
CAN3_A3	CHH10_A8_P	HG2_A6_P
CAN3_A5	CHH10_A9_P	HG2_A9_P
CAN3_A6	CHH10_A10	HG2_A10_P
CAN3_A7	CHH10_A11_P	HG2_A11_P
CAN3_A9_P	*C. h. natalensis*	HG2_A12_P
CAN3_A11	CHN4_A11	HG2_B1_P
CAN4_B12_P	CHN4_A2_P	HG2_B7
CAN4_A11_P	CHN4_A5	HG2_B9_P
CAN4_B5_P	CHN4_A8	HG4_A12_P
CAN4_B6_P	CHN4_A10_P	HG4_A6
CAN4_B7_P	*C.h. pretoriae*	HG4_A9_P
CAN4_B10_P	CHP2_A10_P	HG4_A10_P
F. bocagei	CHP2_A5	
CB1_A12	CHP2_A6	
CB1_A2_P	CHP3_A3_P	
CB1_A4_P	CHP3_A7_P	
CB1_A10_P	CHP3_A10_P	
CB2_A12	CHP3_A11_P	
CB2_A3_P	CHP3_A12	
CB2_A4	CHP3_B6_P	
CB2_A7	*F. mechowi*	
CB2_A10_P	CM6_A12	
	CM6_A2	
	CM6_A6_P	

University of Cape Town

I.4 List of OR allelic pairs Sequences listed on the same row represent allelic variants of the same OR gene. Only sequences in the first column are retained in subsequent analyses as representatives of those particular OR genes. Abbreviations correspond to gene names as per Appendix I.2. The mean % sequence variation between allelic pairs is reported in the last column; alleles that are 99% similar diplay between 3-5 base pair (bp) differences, those that are 98% similar have between 6-10 bp differences, and the 96% similar have between 11-20 bp differences.

List of OR alleles					% sequence similarity
CB2_A7	CM6_A4				98
CHP2_A10_P	CHP3-B11_P				99
CAN3_A6	CDM6_C10				98
CDM6_A5	CDM6_A4				99
CDM6_A8	CDM6_C5				99
CDM8_A3	CA3_A10	CAN4_B11			99
CAN4_B12_P	CDM8_A10_P	CB1_A3_P	CD7b_A9_P		98
CA3_A9	CA3_A12				99
HG2_A11_P	HG4_A3_P				99
HG2_A12_P	HG4_A1_P				99
HG2_B1_P	HG2_B4_P	HG4_A2_P			99
HG2_A4_P	HG2_A7_P				99
HG2_B9_P	HG4_A7_P				99
HG2_A6_P	HG4_A8_P	HG4_A11_P			99
HA1_A10	HA3_A2	HA3_A12			99
GC10_A2_P	GC10_A11_P				99
CD7b_B5_P	CM6_A5_P				99
CHP3_A7_P	CHP3_B5_P	CHN4_A1_P			99
CHN4_A2_P	CHN4_A4_P				99
HA1_A7	HA3_A11	HA3_A1			99
CM6_A6_P	CM6_A9_P				99
BJ4_A4_P	BJ4_A6_P	BS7_A9_P			99
CD7_A2_P	CD7b_A3_P				99
CDM8_A12_P	CDM4_A9_P				99
CDM4_A4_P	CDM4_A6_P				99
CAN3_A5	CB2_A6				98
CHH10_A1	CHH10_A6				99
CHN4_A5	CHN4_A7				99
BJ4_A1	BS7_A2				99
HA1_A3	HA3_A10				99
CAN3_A2	CAN4_B4	CA1_A5			99
GC10_A4	GC10_A8	CHP2_A9			96
CAN4_A11_P	CAN4_A12_P	CA3_B1_P			99
CB2_A10_P	CB2_A9_P				99
CB1_A2_P	CB2_A1_P	CB2_A8_P			99
CB1_A4_P	CB2_A5_P				99
CAN3_A3	CDM4_A2				98
HA3_A9	HA3_A7				99
CAN3_A11	CAN3_A10	CDM8_A6	CM6_A3	CM6_A8	99
CM6_A2	CM6_A7	CD7_A6	CAN4_B9	CAN3_A4	99

I.5 List of functional trans-species polymorphisms, TSPs, and ancient OR loci Sequences listed on the same row represent functional allelic variants found across bathyergid species. Allelic variants that share >99% sequence similarity across different bathyergid species are considered TSP (Klein et al. 1998) and are highlighted in yellow. Abbreviations correspond to gene names as per Appendix I.2.

Functional TSPs and conserved OR loci across bathyergid species					Clade
CB2_A7	CM6_A4				A
CAN3_A6	CDM6_C10				A
CDM8_A3	CA3_A10	CAN4_B11			A
CAN4_B11	CA3_B2				A
CAN3_A5	CB2_A6				D
BJ4_A1	BS7_A2				D
CAN3_A2	CAN4_B4	CA1_A5			D
GC10_A4	GC10_A8	CHP2_A9			D
CAN3_A3	CDM4_A2				D
CAN3_A11	CAN3_A10	CDM8_A6	CM6_A3	CM6_A8	D
CM6_A2	CM6_A7	CD7_A6	CAN4_B9	CAN3_A4	D
CHP2_A9	CHN4_A12				D

I.6 Average sequence similarity (%) between clades A-D Sequence similarity between clades was calculated via the average pairwise distances (based on the number of nucleotide differences) between functional ORs from clades A-D. The result is reported in the table below as an average percentage of between-clades sequence similarity.

	Clade A	Clade B	Clade C
Clade A			
Clade B	66		
Clade C	63	62	
Clade D	68	66	65

University of Cape Town

I.7 Results of the branch-site test of positive selection (CodeML) across the Bathyergidae OR gene tree CodeML is a program from the package PAML that estimates ω = dN/dS across the phylogenetic tree (Yang and Nielsen 2002, Zhang et al. 2005). Two nested models, null (0) and alternative (1), are computed and compared via a likelihood ratio test, LRT. For each model, log likelihood values - lnL_1 for the alternative and lnL_0 for the null models - are used to compute the (LRT); the $2\times(lnL_1\text{-}lnL_0)$ follows a χ^2 curve with degree of freedom of 1 (np1-np0 = 1), from which the P-value is calculated. Subsequently, the Q-value is calculated (Storey 2002, 2003, Storey et al. 2004) as a measure of the false discovery rate due to multiple testing. When the LRT is significant (p-value< 0.05) and the false discovery rate is low (i.e. less than 15%), the posterior probability of sites being under positive selection (dN/dS > 1) is calculated using the Bayes Empirical Bayes (BEB) method (Nielsen and Yang 1998, Yang et al. 2005b) implemented in CodeML. Branches in the bathyergid OR tree were labelled with different numbers by CodeML, indicated in the **Branch** column. Where significant adaptive selection is detected, branches are highlighted in yellow and correspond to the homonymous labelled branches in Figure 3.4 and Appendix III.2; where codon sites under significant positive selection are detected by the BEB method, the initials 'BEB' are indicated in the **Tree** column.

	np0	lnL0	np1	lnL1	branch	LRT	Pvalue	Qvalue	Tree
GEN.B75	239	-12668	240	-12655	75	2.6e+01	3.2e-07	3.8e-05	Branch 75 (PHY) Branch 75 (PDF) BEB
GEN.B34	239	-12668	240	-12658	34	2.0e+01	9.5e-06	5.5e-04	Branch 34 (PHY) Branch 34 (PDF) BEB
GEN.B27	239	-12666	240	-12662	27	8.6e+00	3.4e-03	1.3e-01	Branch 27 (PHY) Branch 27 (PDF) BEB
GEN.B63	239	-12668	240	-12664	63	8.0e+00	4.6e-03	1.3e-01	Branch 63 (PHY) Branch 63 (PDF) BEB
GEN.B54	239	-12668	240	-12665	54	5.1e+00	2.3e-02	5.4e-01	Branch 54 (PHY) Branch 54 (PDF) BEB
GEN.B39	239	-12667	240	-12665	39	4.7e+00	3.1e-02	6.0e-01	Branch 39 (PHY) Branch 39 (PDF) BEB
GEN.B6	239	-12667	240	-12666	6	2.8e+00	9.5e-02	1.0e+00	Branch 6 (PHY) Branch 6 (PDF) BEB
GEN.B44	239	-12667	240	-12665	44	2.6e+00	1.1e-01	1.0e+00	Branch 44 (PHY) Branch 44 (PDF) BEB
GEN.B38	239	-12665	240	-12664	38	1.8e+00	1.8e-01	1.0e+00	Branch 38 (PHY) Branch 38 (PDF) BEB
GEN.B47	239	-12668	240	-12667	47	1.5e+00	2.1e-01	1.0e+00	Branch 47 (PHY) Branch 47 (PDF) BEB
GEN.B5	239	-12667	240	-12667	5	1.3e+00	2.5e-01	1.0e+00	Branch 5 (PHY) Branch 5 (PDF) BEB
GEN.B56	239	-12667	240	-12667	56	1.3e+00	2.5e-01	1.0e+00	Branch 56 (PHY) Branch 56 (PDF) BEB
GEN.B106	239	-12667	240	-12667	106	1.3e+00	2.6e-01	1.0e+00	Branch 106 (PHY) Branch 106 (PDF) BEB

University of Cape Town

GEN.B68	239	-12667	240	-12666	68	1.2e+00	2.6e-01	1.0e+00	Branch 68 (PHY) Branch 68 (PDF) BEB
GEN.B46	239	-12667	240	-12667	46	1.2e+00	2.7e-01	1.0e+00	Branch 46 (PHY) Branch 46 (PDF) BEB
GEN.B100	239	-12668	240	-12667	100	1.2e+00	2.8e-01	1.0e+00	Branch 100 (PHY) Branch 100 (PDF)
GEN.B78	239	-12668	240	-12667	78	1.1e+00	2.9e-01	1.0e+00	Branch 78 (PHY) Branch 78 (PDF) BEB
GEN.B90	239	-12667	240	-12667	90	1.1e+00	3.0e-01	1.0e+00	Branch 90 (PHY) Branch 90 (PDF) BEB
GEN.B92	239	-12668	240	-12668	92	7.2e-01	4.0e-01	1.0e+00	Branch 92 (PHY) Branch 92 (PDF) BEB
GEN.B15	239	-12668	240	-12668	15	3.5e-01	5.5e-01	1.0e+00	Branch 15 (PHY) Branch 15 (PDF) BEB
GEN.B10	239	-12668	240	-12668	10	3.1e-01	5.8e-01	1.0e+00	Branch 10 (PHY) Branch 10 (PDF)
GEN.B72	239	-12668	240	-12668	72	2.2e-01	6.4e-01	1.0e+00	Branch 72 (PHY) Branch 72 (PDF) BEB
GEN.B37	239	-12667	240	-12667	37	1.8e-01	6.7e-01	1.0e+00	Branch 37 (PHY) Branch 37 (PDF) BEB
GEN.B95	239	-12668	240	-12668	95	6.3e-02	8.0e-01	1.0e+00	Branch 95 (PHY) Branch 95 (PDF) BEB
GEN.B29	239	-12668	240	-12668	29	3.7e-02	8.5e-01	1.0e+00	Branch 29 (PHY) Branch 29 (PDF) BEB
GEN.B51	239	-12668	240	-12668	51	8.5e-03	9.3e-01	1.0e+00	Branch 51 (PHY) Branch 51 (PDF)
GEN.B96	239	-12668	240	-12668	96	4.9e-03	9.4e-01	1.0e+00	Branch 96 (PHY) Branch 96 (PDF)
GEN.B111	239	-12668	240	-12668	111	1.9e-04	9.9e-01	1.0e+00	Branch 111 (PHY) Branch 111 (PDF)
GEN.B102	239	-12668	240	-12668	102	2.0e-06	1.0e+00	1.0e+00	Branch 102 (PHY) Branch 102 (PDF)
GEN.B1	239	-12668	240	-12668	1	0.0e+00	1.0e+00	1.0e+00	Branch 1 (PHY) Branch 1 (PDF)
GEN.B101	239	-12668	240	-12668	101	0.0e+00	1.0e+00	1.0e+00	Branch 101 (PHY) Branch 101 (PDF)
GEN.B103	239	-12668	240	-12668	103	0.0e+00	1.0e+00	1.0e+00	Branch 103 (PHY) Branch 103 (PDF)
GEN.B104	239	-12668	240	-12668	104	0.0e+00	1.0e+00	1.0e+00	Branch 104 (PHY) Branch 104 (PDF) BEB
GEN.B105	239	-12668	240	-12668	105	0.0e+00	1.0e+00	1.0e+00	Branch 105 (PHY) Branch 105 (PDF)
GEN.B107	239	-12668	240	-12668	107	0.0e+00	1.0e+00	1.0e+00	Branch 107 (PHY) Branch 107 (PDF)
GEN.B108	239	-12667	240	-12667	108	-4.0e-05	1.0e+00	1.0e+00	Branch 108 (PHY) Branch 108 (PDF) BEB
GEN.B109	239	-12668	240	-12668	109	0.0e+00	1.0e+00	1.0e+00	Branch 109 (PHY) Branch 109 (PDF) BEB
GEN.B11	239	-12668	240	-12668	11	0.0e+00	1.0e+00	1.0e+00	Branch 11 (PHY) Branch 11 (PDF)
GEN.B110	239	-12668	240	-12668	110	0.0e+00	1.0e+00	1.0e+00	Branch 110 (PHY) Branch 110 (PDF)
GEN.B112	239	-12668	240	-12668	112	0.0e+00	1.0e+00	1.0e+00	Branch 112 (PHY)

University of Cape Town

									Branch 112 (PDF)
GEN.B113	239	-12667	240	-12667	113	0.0e+00	1.0e+00	1.0e+00	Branch 113 (PHY) Branch 113 (PDF) BEB
GEN.B114	239	-12668	240	-12668	114	0.0e+00	1.0e+00	1.0e+00	Branch 114 (PHY) Branch 114 (PDF)
GEN.B115	239	-12668	240	-12668	115	0.0e+00	1.0e+00	1.0e+00	Branch 115 (PHY) Branch 115 (PDF)
GEN.B116	239	-12668	240	-12668	116	0.0e+00	1.0e+00	1.0e+00	Branch 116 (PHY) Branch 116 (PDF)
GEN.B12	239	-12668	240	-12668	12	0.0e+00	1.0e+00	1.0e+00	Branch 12 (PHY) Branch 12 (PDF)
GEN.B13	239	-12668	240	-12668	13	0.0e+00	1.0e+00	1.0e+00	Branch 13 (PHY) Branch 13 (PDF)
GEN.B14	239	-12668	240	-12668	14	0.0e+00	1.0e+00	1.0e+00	Branch 14 (PHY) Branch 14 (PDF)
GEN.B16	239	-12668	240	-12668	16	0.0e+00	1.0e+00	1.0e+00	Branch 16 (PHY) Branch 16 (PDF)
GEN.B17	239	-12668	240	-12668	17	0.0e+00	1.0e+00	1.0e+00	Branch 17 (PHY) Branch 17 (PDF)
GEN.B18	239	-12668	240	-12668	18	0.0e+00	1.0e+00	1.0e+00	Branch 18 (PHY) Branch 18 (PDF)
GEN.B19	239	-12668	240	-12668	19	0.0e+00	1.0e+00	1.0e+00	Branch 19 (PHY) Branch 19 (PDF)
GEN.B2	239	-12668	240	-12668	2	0.0e+00	1.0e+00	1.0e+00	Branch 2 (PHY) Branch 2 (PDF)
GEN.B20	239	-12667	240	-12667	20	0.0e+00	1.0e+00	1.0e+00	Branch 20 (PHY) Branch 20 (PDF) BEB
GEN.B21	239	-12668	240	-12668	21	0.0e+00	1.0e+00	1.0e+00	Branch 21 (PHY) Branch 21 (PDF)
GEN.B22	239	-12668	240	-12668	22	0.0e+00	1.0e+00	1.0e+00	Branch 22 (PHY) Branch 22 (PDF)
GEN.B23	239	-12668	240	-12668	23	0.0e+00	1.0e+00	1.0e+00	Branch 23 (PHY) Branch 23 (PDF)
GEN.B24	239	-12668	240	-12668	24	0.0e+00	1.0e+00	1.0e+00	Branch 24 (PHY) Branch 24 (PDF)
GEN.B25	239	-12668	240	-12668	25	-2.0e-06	1.0e+00	1.0e+00	Branch 25 (PHY) Branch 25 (PDF)
GEN.B26	239	-12668	240	-12668	26	0.0e+00	1.0e+00	1.0e+00	Branch 26 (PHY) Branch 26 (PDF)
GEN.B28	239	-12668	240	-12668	28	0.0e+00	1.0e+00	1.0e+00	Branch 28 (PHY) Branch 28 (PDF)
GEN.B3	239	-12668	240	-12668	3	-1.0e-05	1.0e+00	1.0e+00	Branch 3 (PHY) Branch 3 (PDF)
GEN.B30	239	-12668	240	-12668	30	0.0e+00	1.0e+00	1.0e+00	Branch 30 (PHY) Branch 30 (PDF) BEB
GEN.B31	239	-12668	240	-12668	31	0.0e+00	1.0e+00	1.0e+00	Branch 31 (PHY) Branch 31 (PDF)
GEN.B32	239	-12668	240	-12668	32	-6.0e-06	1.0e+00	1.0e+00	Branch 32 (PHY) Branch 32 (PDF)
GEN.B33	239	-12668	240	-12668	33	0.0e+00	1.0e+00	1.0e+00	Branch 33 (PHY) Branch 33 (PDF)
GEN.B35	239	-12668	240	-12668	35	0.0e+00	1.0e+00	1.0e+00	Branch 35 (PHY) Branch 35 (PDF)
GEN.B36	239	-12668	240	-12668	36	0.0e+00	1.0e+00	1.0e+00	Branch 36 (PHY) Branch 36 (PDF)
GEN.B4	239	-12668	240	-12668	4	-2.0e-06	1.0e+00	1.0e+00	Branch 4 (PHY) Branch 4 (PDF)
GEN.B40	239	-12668	240	-12668	40	0.0e+00	1.0e+00	1.0e+00	Branch 40 (PHY) Branch 40 (PDF)
GEN.B41	239	-12668	240	-12668	41	0.0e+00	1.0e+00	1.0e+00	Branch 41 (PHY) Branch 41 (PDF)

University of Cape Town

GEN.B42	239	-12667	240	-12667	42	0.0e+00	1.0e+00	1.0e+00	Branch 42 (PHY) Branch 42 (PDF) BEB
GEN.B43	239	-12668	240	-12668	43	0.0e+00	1.0e+00	1.0e+00	Branch 43 (PHY) Branch 43 (PDF) BEB
GEN.B45	239	-12668	240	-12668	45	0.0e+00	1.0e+00	1.0e+00	Branch 45 (PHY) Branch 45 (PDF)
GEN.B48	239	-12668	240	-12668	48	0.0e+00	1.0e+00	1.0e+00	Branch 48 (PHY) Branch 48 (PDF) BEB
GEN.B49	239	-12668	240	-12668	49	0.0e+00	1.0e+00	1.0e+00	Branch 49 (PHY) Branch 49 (PDF) BEB
GEN.B50	239	-12668	240	-12668	50	-4.7e-02	1.0e+00	1.0e+00	Branch 50 (PHY) Branch 50 (PDF)
GEN.B52	239	-12668	240	-12668	52	0.0e+00	1.0e+00	1.0e+00	Branch 52 (PHY) Branch 52 (PDF)
GEN.B53	239	-12668	240	-12668	53	0.0e+00	1.0e+00	1.0e+00	Branch 53 (PHY) Branch 53 (PDF)
GEN.B55	239	-12668	240	-12668	55	0.0e+00	1.0e+00	1.0e+00	Branch 55 (PHY) Branch 55 (PDF)
GEN.B57	239	-12668	240	-12668	57	0.0e+00	1.0e+00	1.0e+00	Branch 57 (PHY) Branch 57 (PDF)
GEN.B58	239	-12668	240	-12668	58	-3.8e-05	1.0e+00	1.0e+00	Branch 58 (PHY) Branch 58 (PDF)
GEN.B59	239	-12668	240	-12668	59	0.0e+00	1.0e+00	1.0e+00	Branch 59 (PHY) Branch 59 (PDF)
GEN.B60	239	-12668	240	-12668	60	-2.0e-06	1.0e+00	1.0e+00	Branch 60 (PHY) Branch 60 (PDF)
GEN.B61	239	-12668	240	-12668	61	0.0e+00	1.0e+00	1.0e+00	Branch 61 (PHY) Branch 61 (PDF) BEB
GEN.B62	239	-12668	240	-12668	62	0.0e+00	1.0e+00	1.0e+00	Branch 62 (PHY) Branch 62 (PDF)
GEN.B64	239	-12668	240	-12668	64	0.0e+00	1.0e+00	1.0e+00	Branch 64 (PHY) Branch 64 (PDF) BEB
GEN.B65	239	-12668	240	-12668	65	0.0e+00	1.0e+00	1.0e+00	Branch 65 (PHY) Branch 65 (PDF)
GEN.B66	239	-12668	240	-12668	66	0.0e+00	1.0e+00	1.0e+00	Branch 66 (PHY) Branch 66 (PDF)
GEN.B67	239	-12668	240	-12668	67	0.0e+00	1.0e+00	1.0e+00	Branch 67 (PHY) Branch 67 (PDF)
GEN.B69	239	-12668	240	-12668	69	0.0e+00	1.0e+00	1.0e+00	Branch 69 (PHY) Branch 69 (PDF)
GEN.B7	239	-12668	240	-12668	7	0.0e+00	1.0e+00	1.0e+00	Branch 7 (PHY) Branch 7 (PDF)
GEN.B70	239	-12668	240	-12668	70	-4.0e-06	1.0e+00	1.0e+00	Branch 70 (PHY) Branch 70 (PDF)
GEN.B71	239	-12668	240	-12668	71	0.0e+00	1.0e+00	1.0e+00	Branch 71 (PHY) Branch 71 (PDF)
GEN.B73	239	-12668	240	-12668	73	0.0e+00	1.0e+00	1.0e+00	Branch 73 (PHY) Branch 73 (PDF) BEB
GEN.B74	239	-12668	240	-12668	74	0.0e+00	1.0e+00	1.0e+00	Branch 74 (PHY) Branch 74 (PDF) BEB
GEN.B76	239	-12668	240	-12668	76	0.0e+00	1.0e+00	1.0e+00	Branch 76 (PHY) Branch 76 (PDF) BEB
GEN.B77	239	-12668	240	-12668	77	0.0e+00	1.0e+00	1.0e+00	Branch 77 (PHY) Branch 77 (PDF)
GEN.B79	239	-12668	240	-12668	79	0.0e+00	1.0e+00	1.0e+00	Branch 79 (PHY) Branch 79 (PDF)

University of Cape Town

GEN.B8	239	-12668	240	-12668	8	0.0e+00	1.0e+00	1.0e+00	Branch 8 (PHY) Branch 8 (PDF)
GEN.B80	239	-12668	240	-12668	80	0.0e+00	1.0e+00	1.0e+00	Branch 80 (PHY) Branch 80 (PDF)
GEN.B81	239	-12668	240	-12668	81	0.0e+00	1.0e+00	1.0e+00	Branch 81 (PHY) Branch 81 (PDF)
GEN.B82	239	-12668	240	-12668	82	0.0e+00	1.0e+00	1.0e+00	Branch 82 (PHY) Branch 82 (PDF)
GEN.B83	239	-12668	240	-12668	83	0.0e+00	1.0e+00	1.0e+00	Branch 83 (PHY) Branch 83 (PDF)
GEN.B84	239	-12668	240	-12668	84	0.0e+00	1.0e+00	1.0e+00	Branch 84 (PHY) Branch 84 (PDF)
GEN.B85	239	-12667	240	-12667	85	0.0e+00	1.0e+00	1.0e+00	Branch 85 (PHY) Branch 85 (PDF) BEB
GEN.B86	239	-12668	240	-12668	86	0.0e+00	1.0e+00	1.0e+00	Branch 86 (PHY) Branch 86 (PDF)
GEN.B87	239	-12668	240	-12668	87	0.0e+00	1.0e+00	1.0e+00	Branch 87 (PHY) Branch 87 (PDF)
GEN.B88	239	-12668	240	-12668	88	0.0e+00	1.0e+00	1.0e+00	Branch 88 (PHY) Branch 88 (PDF)
GEN.B89	239	-12668	240	-12668	89	0.0e+00	1.0e+00	1.0e+00	Branch 89 (PHY) Branch 89 (PDF) BEB
GEN.B9	239	-12668	240	-12668	9	0.0e+00	1.0e+00	1.0e+00	Branch 9 (PHY) Branch 9 (PDF)
GEN.B91	239	-12668	240	-12668	91	0.0e+00	1.0e+00	1.0e+00	Branch 91 (PHY) Branch 91 (PDF)
GEN.B93	239	-12668	240	-12668	93	0.0e+00	1.0e+00	1.0e+00	Branch 93 (PHY) Branch 93 (PDF)
GEN.B94	239	-12668	240	-12668	94	0.0e+00	1.0e+00	1.0e+00	Branch 94 (PHY) Branch 94 (PDF)
GEN.B97	239	-12668	240	-12668	97	0.0e+00	1.0e+00	1.0e+00	Branch 97 (PHY) Branch 97 (PDF) BEB
GEN.B98	239	-12668	240	-12668	98	0.0e+00	1.0e+00	1.0e+00	Branch 98 (PHY) Branch 98 (PDF) BEB
GEN.B99	239	-12668	240	-12668	99	0.0e+00	1.0e+00	1.0e+00	Branch 99 (PHY) Branch 99 (PDF)

University of Cape Town

I.8 List of mammalian species and OR 1/3/7 information used for the Ecogroup analysis OR 1/3/7 information for non-bathyergid species is reported as per Hayden et al. (2010); OR functional gene and pseudogene numbers are indicated under the **Pseudo OR1/3/7** and **OR 1/3/7** columns, respectively; OR pseudogene proportions (%) are indicated in the **% pseudo OR 1/3/7** column; species are classified into five 'ecogroups' according to the environmental niche that species inhabit: Aquatic, Semi-aquatic, Terrestrial, Volant (bats) and Subterranean (bathyergids).

Superorder	Order	Family	Genus	Species	Abbreviation	Common name	Ecogroup	Pseudo OR 1/3/7	OR 1/3/7	% pseudo OR 1/3/7
Afrotheria	Afrosoricida	Tenrecidae	Echinops	Echinops telfairi	EteG	Small madagascar hedgehog	Terrestrial	55	84	40%
Afrotheria	Hyracoidea	Procaviidae	Procavia	Procavia capensis	PcaG	Rock hyrax	Terrestrial	46	35	57%
Afrotheria	Proboscidea	Elephantidae	Loxodonta	Loxodonta africana	LafG	African savanna elephant	Terrestrial	208	207	50%
Afrotheria	Sirenia	Trichechidae	Trichechus	Trichechus manatus	Tma	Caribbean manatee	Aquatic	10	7	59%
Afrotheria	Xenarthra	Dasypodidae	Dasypus	Dasypus novemcinctus	DnoG	Nine-banded armadillo	Terrestrial	372	190	66%
Euarchontoglires	Lagomorpha	Ochotonidae	Ochotona	Ochotona princeps	OprG	American pika	Terrestrial	96	87	52%
Euarchontoglires	Lagomorpha	Leporidae	Oryctolagus	Oryctolagus cuniculus	OcuG	Rabbit	Terrestrial	89	94	49%
Euarchontoglires	Primates	Hominidae	Gorilla	Gorilla gorilla	GgoG	Western lowland gorilla	Terrestrial	54	32	63%
Euarchontoglires	Primates	Hominidae	Homo	Homo sapiens	HsaG	Human	Terrestrial	88	69	56%
Euarchontoglires	Primates	Cercopithecidae	Macaca	Macaca mulatta	MmaG	Rhesus monkey	Terrestrial	34	31	52%
Euarchontoglires	Primates	Cheirogaleidae	Microcebus	Microcebus murinus	MmuG	Grey mouse lemur	Terrestrial	43	62	41%
Euarchontoglires	Primates	Galagidae	Otolemur	Otolemur garnettii	OgaG	Northern Greater Galago	Terrestrial	36	34	51%
Euarchontoglires	Primates	Hominidae	Pan	Pan troglodytes	PtrG	Chimpanzee	Terrestrial	103	65	61%
Euarchontoglires	Primates	Hominidae	Pongo	Pongo pygmaeus	PpyG	Orangutan	Terrestrial	110	69	61%
Euarchontoglires	Primates	Tarsiidae	Tarsius	Tarsius syrichta	TsyG	Philippine Tarsier	Terrestrial	45	10	82%
Euarchontoglires	Rodentia	Caviidae	Cavia	Cavia porcellus	CpoG	Domestic guinea pig	Terrestrial	88	62	59%
Euarchontoglires	Rodentia	Heteromyidae	Dipodomys	Dipodomys ordii	DorG	Ord's Kangaroo Rat	Terrestrial	30	54	36%
Euarchontoglires	Rodentia	Muridae	Mus	Mus musculus	MusG	House mouse	Terrestrial	13	132	9%
Euarchontoglires	Rodentia	Muridae	Rattus	Rattus norvegicus	RnoG	Norway rat	Terrestrial	46	195	19%
Euarchontoglires	Rodentia	Sciuridae	Spermophilus	Spermophilus tridecemlineatus	StrG	Thirteen-lined ground squirrel	Terrestrial	167	126	57%

University of Cape Town

Euarchontoglires	Rodentia	Bathyergidae	Bathyergus	Bathyergus janetta	BJ	Namaqua dune mole-rat	Subterranean	3	2	60%
Euarchontoglires	Rodentia	Bathyergidae	Bathyergus	Bathyergus suillus	BS	Cape dune mole-rat	Subterranean	5	4	56%
Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Fukomys amatus	CA	Zambian mole-rat	Subterranean	5	6	45%
Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Fukomys anselli	CAN	Ansell's mole-rat	Subterranean	7	8	47%
Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Fukomys bocagei	CB	Bocage's mole-rat	Subterranean	6	5	55%
Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Fukomys damarensis	CDM	Damaraland mole-rat	Subterranean	5	7	42%
Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Fukomys darlingi	CD	Mashona mole-rat	Subterranean	9	1	90%
Euarchontoglires	Rodentia	Bathyergidae	Cryptomys	Cryptomys hottentotus hottentotus	CHH	Common mole-rat	Subterranean	4	6	40%
Euarchontoglires	Rodentia	Bathyergidae	Cryptomys	Cryptomys hottentotus natalensis	CHN	Natal mole-rat	Subterranean	3	3	50%
Euarchontoglires	Rodentia	Bathyergidae	Cryptomys	Cryptomys hottentotus pretoriae	CHP	Highveld mole-rat	Subterranean	6	4	60%
Euarchontoglires	Rodentia	Bathyergidae	Fukomys	Fukomys mechowi	CM	Mechow's mole-rat	Subterranean	2	4	33%
Euarchontoglires	Rodentia	Bathyergidae	Georichus	Georychus capensis	GC	Cape mole-rat	Subterranean	4	5	44%
Euarchontoglires	Rodentia	Bathyergidae	Heliophobius	Heliophobius argentocinereus	HA	Silvery mole-rat	Subterranean	3	7	30%
Euarchontoglires	Rodentia	Bathyergidae	Heterocephalus	Heterocephalus glaber	HG	Naked mole-rat	Subterranean	13	2	87%
Euarchontoglires	Scandentia	Tupaiidae	Tupaia	Tupaia belangeri	TbeG	Northern tree shrew	Terrestrial	381	191	67%
Laurasiatheria	Carnivora	Mustelidae	Aonyx	Aonyx cinerea	Aci	Oriental small-clawed otter	Semi Aquatic	2	13	13%
Laurasiatheria	Carnivora	Otariidae	Arctocephalus	Arctocephalus forsteri	Afo	New Zealand fur seal	Semi Aquatic	5	10	33%
Laurasiatheria	Carnivora	Canidae	Canis	Canis familiaris	CfaG	Dog (Boxer)	Terrestrial	21	78	21%
Laurasiatheria	Carnivora	Mustelidae	Enhydra	Enhydra lutris	Elu	Sea otter	Aquatic	1	11	8%
Laurasiatheria	Carnivora	Felidae	Felis	Felis catus	FcaG	Cat	Terrestrial	54	66	45%
Laurasiatheria	Cetacea	Delphinidae	Globicephala	Globicephala sp.	Gsp	Pilot whale	Aquatic	1	0	100%
Laurasiatheria	Cetacea	Balaenopteridae	Megaptera	Megaptera novaeangliae	Mno	Humpback whale	Aquatic	5	4	56%
Laurasiatheria	Cetacea	Delphinidae	Tursiops	Tursiops truncatus	TtrG	Bottlenose Dolphin	Aquatic	3	2	60%
Laurasiatheria	Cetartiodactyla	Bovidae	Bos	Bos taurus	BtaG	Cattle	Terrestrial	99	220	31%
Laurasiatheria	Cetartiodactyla	Hippopotamidae	Hippopotamus	Hippopotamus amphibius	Ham	Hippopotamus	Semi Aquatic	4	15	21%

University of Cape Town

Laurasiatheria	Cetartiodactyla	Camelidae	Lama	Lama pacos	LpaG	Alpaca	Terrestrial	16	29	36%
Laurasiatheria	Chiroptera	Phyllostomidae	Anoura	Anoura geoffroyi	Age	Geoffroy's tailless bat	Volant	34	68	33%
Laurasiatheria	Chiroptera	Phyllostomidae	Artibeus	Artibeus jamaicensis	Aja	Jamaican fruit-eating bat	Volant	12	18	40%
Laurasiatheria	Chiroptera	Pteropodidae	Cynopterus	Cynopterus sphinx	Csp	Indian short-nosed fruit bat	Volant	13	61	18%
Laurasiatheria	Chiroptera	Emballonuridae	Emballonura	Emballonura atrata	Eat	Peters's sheath-tailed bat	Volant	24	43	36%
Laurasiatheria	Chiroptera	Molossidae	Eumops	Eumops auripendulus	Eau	Black bonneted bat	Volant	21	35	38%
Laurasiatheria	Chiroptera	Vespertilionidae	Myotis	Myotis lucifugus	Mlu	Little brown bat	Volant	9	38	19%
Laurasiatheria	Chiroptera	Pteropodidae	Nyctimene	Nyctimene albiventer	Nal	Common tube-nosed fruit bat	Volant	7	12	37%
Laurasiatheria	Chiroptera	Mormoopidae	Pteronotus	Pteronotus parnellii	Ppa	Parnell's mustached bat	Volant	20	31	39%
Laurasiatheria	Chiroptera	Pteropodidae	Pteropus	Pteropus rayneri	Pra	Solomons flying fox	Volant	9	26	26%
Laurasiatheria	Chiroptera	Pteropodidae	Pteropus	Pteropus vampyrus	PvaG	Large Flying Fox	Volant	87	83	51%
Laurasiatheria	Chiroptera	Rhinolophidae	Rhinolophus	Rhinolophus hipposideros	Rhi	Lesser horseshoe bat	Volant	1	10	9%
Laurasiatheria	Chiroptera	Pteropodidae	Rousettus	Rousettus lanosus	Rla	Long-haired rousette	Volant	9	29	24%
Laurasiatheria	Insectivora	Erinaceidae	Erinaceus	Erinaceus europaeus	EeuG	Western european hedgehog	Terrestrial	35	35	50%
Laurasiatheria	Insectivora	Soricidae	Sorex	Sorex araneus	SarG	European shrew	Terrestrial	197	167	54%
Laurasiatheria	Perissodactyla	Equidae	Equus	Equus caballus	EcaG	Horse	Terrestrial	145	175	45%
Marsupialia	Didelphimorphia	Didelphidae	Monodelphis	Monodelphis domestica	MdoG	Gray short-tailed opossum	Terrestrial	11	83	12%
Monotremata	Monotremata	Ornithorhynchidae	Ornithorhynchus	Ornithorhynchus anatinus	OanG	Platypus	Semi Aquatic	23	39	37%

University of Cape Town

I.9 Sequence annotation based on sociality. Abbreviations correspond to gene names as per Appendix I.2.

ID	Sociality
BJ4_A12	Solitary
BJ4_A1	Solitary
BJ4_A3_P	Solitary
BJ4_A4_P	Solitary
BJ4_A5_P	Solitary
BS7_A1_P	Solitary
BS7_A3_P	Solitary
BS7_A4	Solitary
BS7_A5	Solitary
BS7_A6_P	Solitary
BS7_A7_P	Solitary
BS7_A10	Solitary
CA1_A12_P	Social
CA1_A3_P	Social
CA1_A8	Social
CA3_B3_P	Social
CA3_A8_P	Social
CA3_A9	Social
CA3_A11	Social
CA3_B4	Social
CAN3_A2	Social
CAN3_A3	Social
CAN3_A5	Social
CAN3_A6	Social
CAN3_A7	social
CAN3_A9_P	social
CAN3_A11	social
CAN4_B12_P	social
CAN4_A11_P	social
CAN4_B5_P	social
CAN4_B6_P	social
CAN4_B7_P	social
CAN4_B10_P	social
CB1_A12	social
CB1_A2_P	social
CB1_A4_P	social
CB1_A10_P	social
CB2_A12	social
CB1_A3_P	social
CB2_A4	social
CB2_A7	social
CB2_A10_P	social
CD7_A8_P	social
CD7_A2_P	social
CD7_A7_P	social
CD7b_B5_P	social
CD7b_A6_P	social
CD7b_A11_P	social
CD7b_A12_P	social
CD7b_B3_P	social
CDM4_A3_P	social
CDM4_A4_P	social
CDM6_A5	social
CDM6_A7	social
CDM6_A8	social
CDM8_A12_P	social
CDM8_A3	social
CDM8_A9_P	social
CHH10_A12	social
CHH10_A1	social
CHH10_A2	social
CHH10_A3_P	social
CHH10_A4	social
CHH10_A7	social
CHH10_A8_P	social
CHH10_A9_P	social
CHH10_A10	social
CHH10_A11_P	social
CHN4_A11	social
CHN4_A2_P	social
CHN4_A5	social
CHN4_A8	social
CHN4_A10_P	social
CHP2_A10_P	social

University of Cape Town

CHP2_A5	Social
CHP2_A6	Social
CHP3_A3_P	Social
CHP3_A7_P	Social
CHP3_A10_P	Social
CHP3_A11_P	Social
CHP3_A12	Social
CHP3_B6_P	Social
CM6_A12	Social
CM6_A2	Social
CM6_A6_P	Social
GC8_A11	Solitary
GC8_A3	Solitary
GC8_A6_P	Solitary
GC10_A12_P	Solitary

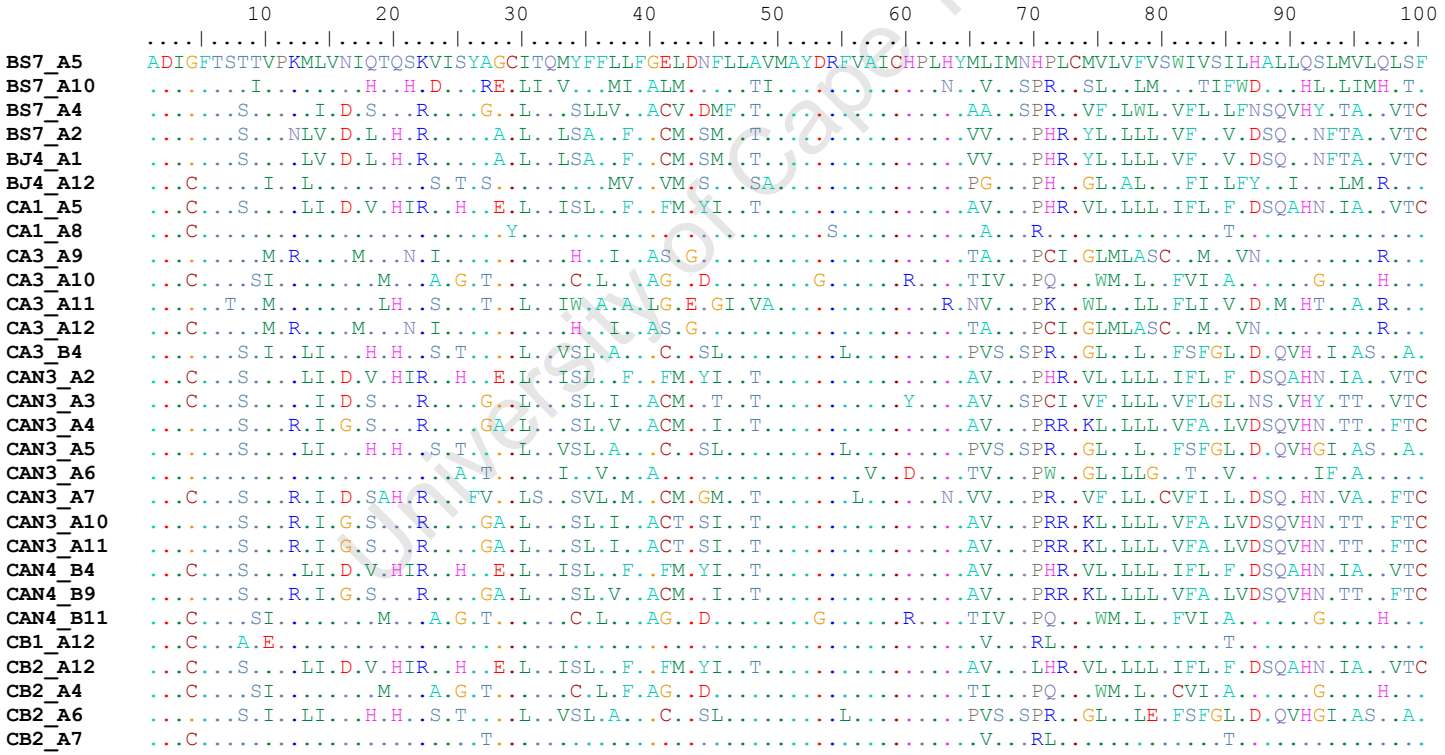
GC10_A2_P	solitary
GC10_A4	solitary
GC10_A5	solitary
GC10_A9	solitary
GC10_A10_P	solitary
HA1_A10	solitary
HA1_A3	solitary
HA1_A4_P	solitary
HA1_A6_P	solitary
HA1_A7	solitary
HA1_A8	solitary
HA1_A9_P	solitary
HA3_A3	solitary
HA3_A4	solitary
HA3_A9	solitary

HG2_B11_P	social
HG2_A2_P	social
HG2_A4_P	social
HG2_A6_P	social
HG2_A9_P	social
HG2_A10_P	social
HG2_A11_P	social
HG2_A12_P	social
HG2_B1_P	social
HG2_B7	social
HG2_B9_P	social
HG4_A12_P	social
HG4_A6	social
HG4_A9_P	social
HG4_A10_P	social

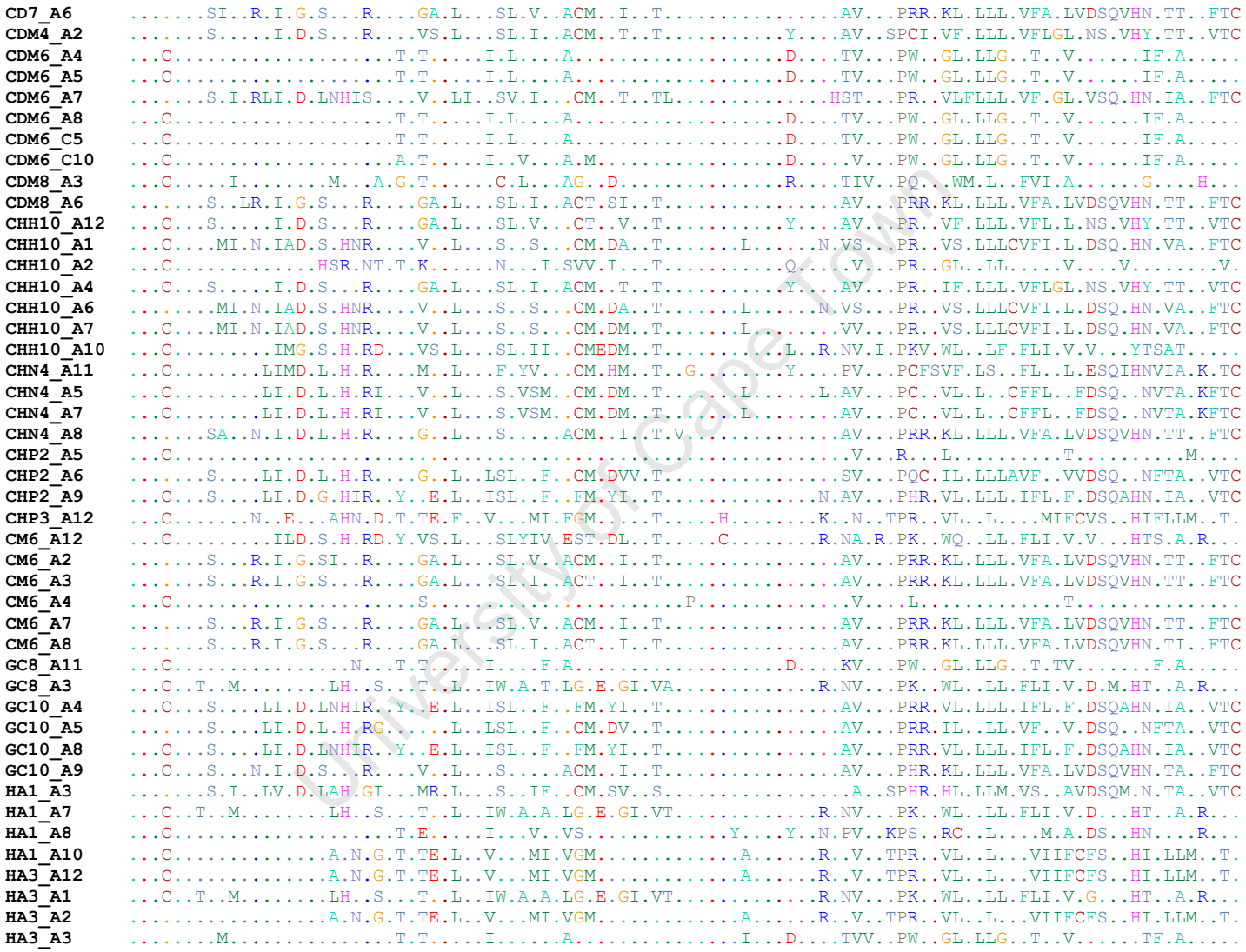
University of Cape Town

Appendix II: Alignments

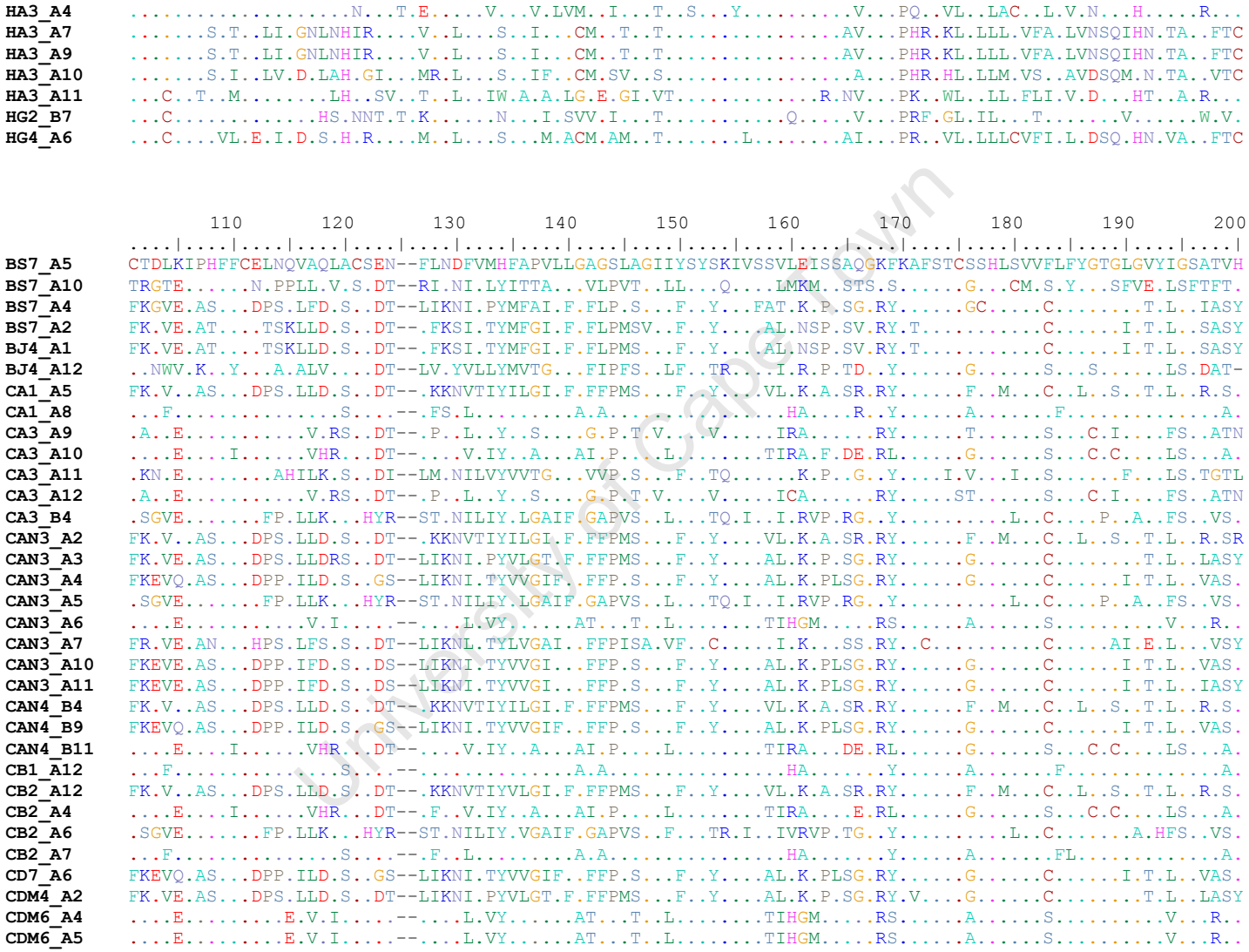
II.1 Protein alignment of the putatively functional bathyergid OR genes Abbreviations correspond to gene names as per Appendix I.2.



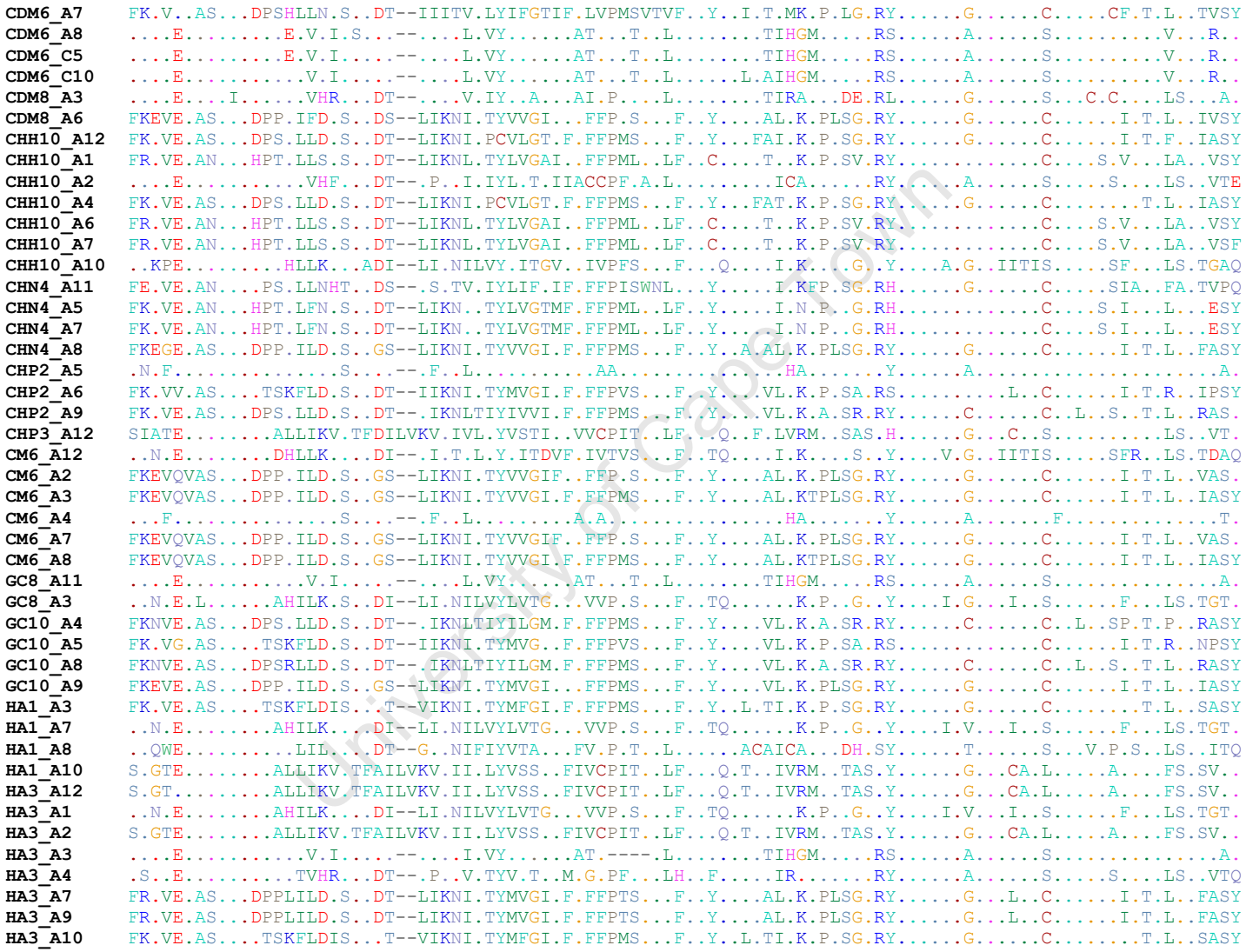
University of Cape Town



University of Cape Town



University of Cape Town



University of Cape Town

```
HA3_A11     ..N.E........AHILK....DI--LI.NILVYLVTG...VVP.S...F..TQ......K.P..G..Y....I.V...I..S......F...LS.TGT.
HG2_B7      .I..E...........VHC...DT--.P..I.IYL.T..IACCPF.A.L.........ICA......RY......V......S.....S....LS..ATE
HG4_A6      FK.VE.AN...HPS.LLR.S.YDT--LIKSL.TYLVGAI..FFPML..LF..C.....I.K...SG.RC.............C.....AI...L..SVSY
```

```
                  210
            ....|....|....|
BS7_A5      SSHSSAKASVMYTVV
BS7_A10     ..QA.MIT.......
BS7_A4      .PKKGMV..L.....
BS7_A2      .PRKGMV..L.....
BJ4_A1      .PRKGMV..L.....
BJ4_A12     ..WRGMV........
CA1_A5      TPRKGMV..L.....
CA1_A8      ......T........
CA3_A9      .....GT........
CA3_A10     N...T.A........
CA3_A11     ..RKN.V........
CA3_A12     .....GT........
CA3_B4      .PKKDMVT.......
CAN3_A2     TPRKGMV..L.....
CAN3_A3     .PRKGMV..L.....
CAN3_A4     .PRKGMV..L.....
CAN3_A5     .PKKDMVT.......
CAN3_A6     ....T.A........
CAN3_A7     .PRKGMV..L.....
CAN3_A10    .PRKGMV..L.....
CAN3_A11    .PRKGMV..L.....
CAN4_B4     TPRKGMV..L.....
CAN4_B9     .PRKGMV..L.....
CAN4_B11    N...T.A........
CB1_A12     ......T........
CB2_A12     TPRKGMV..L.....
CB2_A4      N...T.A........
CB2_A6      .PKKDMVT.......
CB2_A7      ......T........
CD7_A6      .PRKGMV..L.....
CDM4_A2     .PRKGMV..L.....
CDM6_A4     ....T.A........
CDM6_A5     ....T.A........
CDM6_A7     .PSTGEV..L.....
CDM6_A8     ....T.A........
CDM6_C5     ....T.A........
CDM6_C10    ....T.A........
CDM8_A3     N...T.A........
```

University of Cape Town

```
CDM8_A6      .PRKGMV..L.....
CHH10_A12    .PRKGMV..L.....
CHH10_A1     .PRKGMV..L.....
CHH10_A2     N...V.R........
CHH10_A4     .PRKGMV..L.....
CHH10_A6     .PRKGMV..L.....
CHH10_A7     .PRKGMV..L.....
CHH10_A10    .PRK..I........
CHN4_A11     .PRKCMV..L.....
CHN4_A5      ..R.GVV..L.....
CHN4_A7      ..R.GVV..L.....
CHN4_A8      .PRKGMV..L.....
CHP2_A5      ......R........
CHP2_A6      ..RKGMV..L.....
CHP2_A9      TPRKGMV..L.....
CHP3_A12     ..RE.SI........
CM6_A12      .PRK..IT.......
CM6_A2       .PRKGMV..L.....
CM6_A3       .PRKGMV..L.....
CM6_A4       ......T........
CM6_A7       .PRKGMV..L.....
CM6_A8       .PRKGMV..L.....
GC8_A11      G...T.V........
GC8_A3       ..RK..V........
GC10_A4      TPRKGMV..L.....
GC10_A5      .PKKGML..L.....
GC10_A8      TPRKGMV..L.....
GC10_A9      .PRKGMV..L.....
HA1_A3       ..RKGMV..L.....
HA1_A7       ..RK..V........
HA1_A8       N...RST........
HA1_A10      ...E.SV........
HA3_A12      ...E.SV........
HA3_A1       ..RK..V........
HA3_A2       ...E.SV........
HA3_A3       ....T.V........
HA3_A4       N.YATET........
HA3_A7       .PRKGMV..L.....
HA3_A9       .PRKGMV..L.....
HA3_A10      ..RKGMV..L.....
HA3_A11      ..RK..V........
HG2_B7       N...V.T...V....
HG4_A6       .PRKGMVL.L.....
```

University of Cape Town

II.2 Odorant binding sites across clades A-D Only those amino-acid residues involved in odorant binding are reported below, following Man et al. (2004) and Katada et al. (2005); abbreviations correspond to gene names as per Appendix I.2.

Clade A

```
                  10         20
          ....|....| ....|....| ...
BS7_A5    MFLGENFILS FLVALLSLFY TGV
CA1_A8    .......T.. .......... ...
CA3_A9    ..IAS..... ......P... .I.
CAN3_A6   IV.A...V.I .......... ...
CB2_A12   I...FYIFQN .PTL.FPM.L .ST
CB2_A4    .L.AGD.A.G I.....P... .C.
CB2_A7    .......T.. ........L. ...
CDM6_A5   I..A...V.I .......... ...
CDM6_A8   I..A...V.I .......... ...
CDM8_A3   .L.AGD.A.G I.....P... .C.
CHH10_A2  ...SVI.V.. ....IIPF.. .S.
CHP2_A5   .......T.. ......A... ...
GC8_A11   I..A...V.. .......... ...
HA1_A8    I.VVS..A.N ..FT..P... .P.
HA3_A3    I..A...V.T ......I.GL VYS
HA3_A4    V.VVMI.V.. .....MPF.. .S.
HG2_B7    ...SVI.... .....IPF.. .S.
```

Clade B

```
                  10         20
          ....|....| ....|....| ...
CHP3_A12  VFMFGNFFLI FLIYTIVCSL YGL
HA1_A10   ...V...... ....SV..L. ...
BS7_A10   ...AL....L ..VTLLPVYY TSE
```

Clade C

```
                   10         20
           ....|....| ....|....| ...
CA3_A11    IALLGNGVMT FLLVLLPLFY TGV
CHH10_A10  MFIGCDM.L. ...IV..F.. .S.
CM6_A12    MYIESDL.L. ...IVFTV.. .S.
GC8_A3     .......... .......... ...
HA1_A7     ........L. .......... ...
```

University of Cape Town

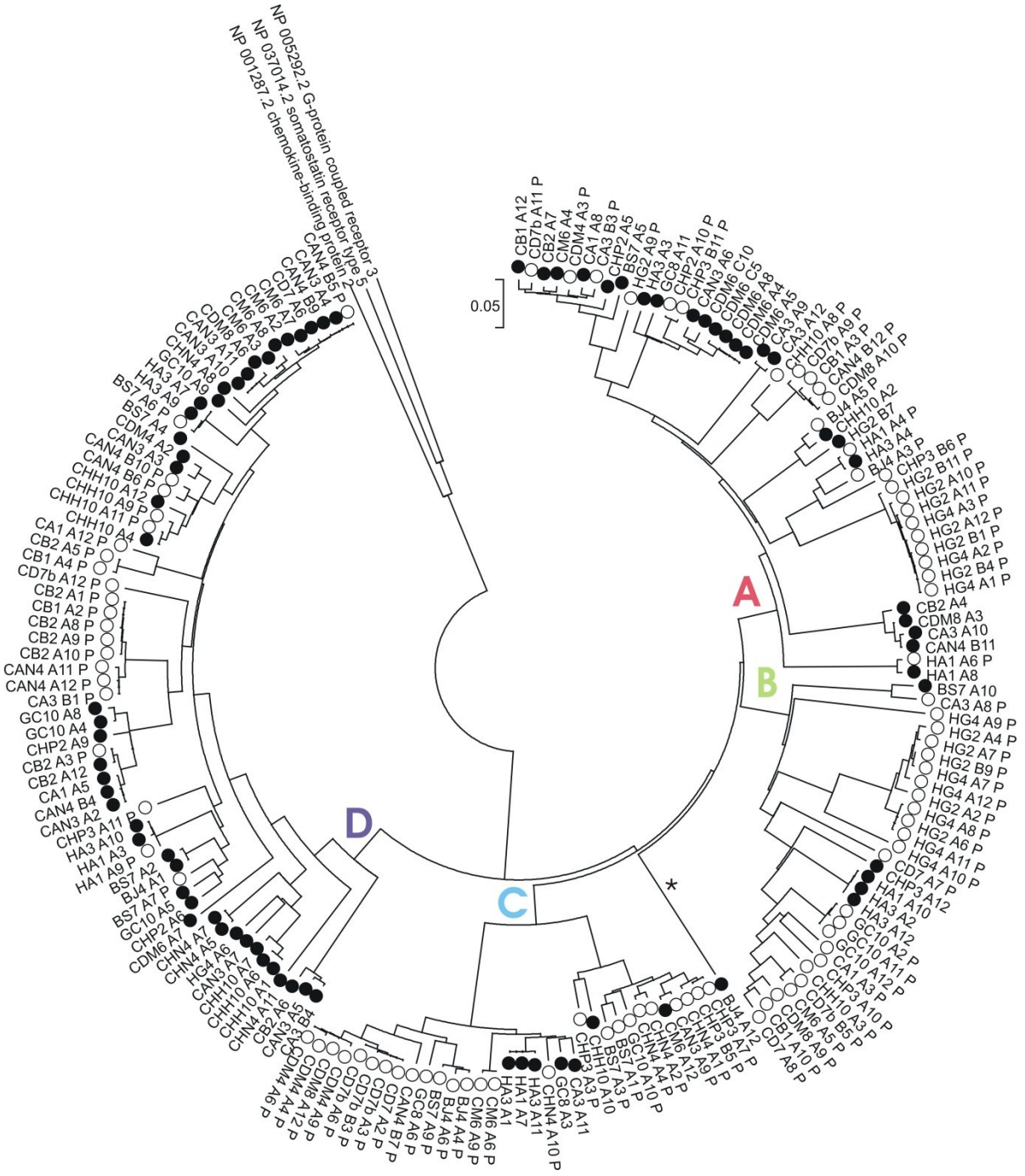
Clade D

```
                    10         20
             ....|....|  ....|....| ...
BJ4_A1       LFLGCSMVQN FTVFLFPMFY TGT
BS7_A4       MLVA.D.L.Y .P.....L.. ...
CA3_B4       V.A...LL.S .FLLI..V.. P.V
CAN3_A2      I...FYIF.. .PTL.....L .S.
CAN3_A3      M.IA.NTLLY .P.L...... ...
CAN3_A5      V.A...LL.G .FLLI..V.. P.V
CAN3_A7      ML...G.L.. .P.VIL.I.. .AE
CAN3_A11     M.IA..IL.. .P.V.L.L.. ...
CB2_A12      I...FYIF.. .PTL.....L .S.
CDM6_A7      M.I..NTL.. .P..I..... .C.
CHH10_A12    M.V..NVLLY .P.L...... ...
CHH10_A1     M.S..DAL.. .P.VIL.... S.V
CHH10_A4     M.IA.NTLLY .P.L...... ...
CHH10_A7     M.S..D.L.. .P.VIL.... S.V
CHN4_A11     MYV..H.L.. .P.II..I.. .SV
CHN4_A5      MVS..D.I.. .P.VM..... S.V
CHN4_A8      M..A.NIL.. .P.V...... ...
CHP2_A6      .....DV... ...V...V.. ...
CM6_A2       M.VA.NIL.. .P.VFL.L.. ...
GC10_A4      I...FYIF.. .PTL.....L .S.
GC10_A5      .....DV... ...V...V.. ...
GC10_A9      M..A.NIL.. .P.V.L.... ...
HA1_A3       M.I...VA.. .......... ...
HA3_A9       M.I..NTL.. .P.V...T.. ...
HG4_A6       M..A.A.L.. .P.VIL.... .AV
```

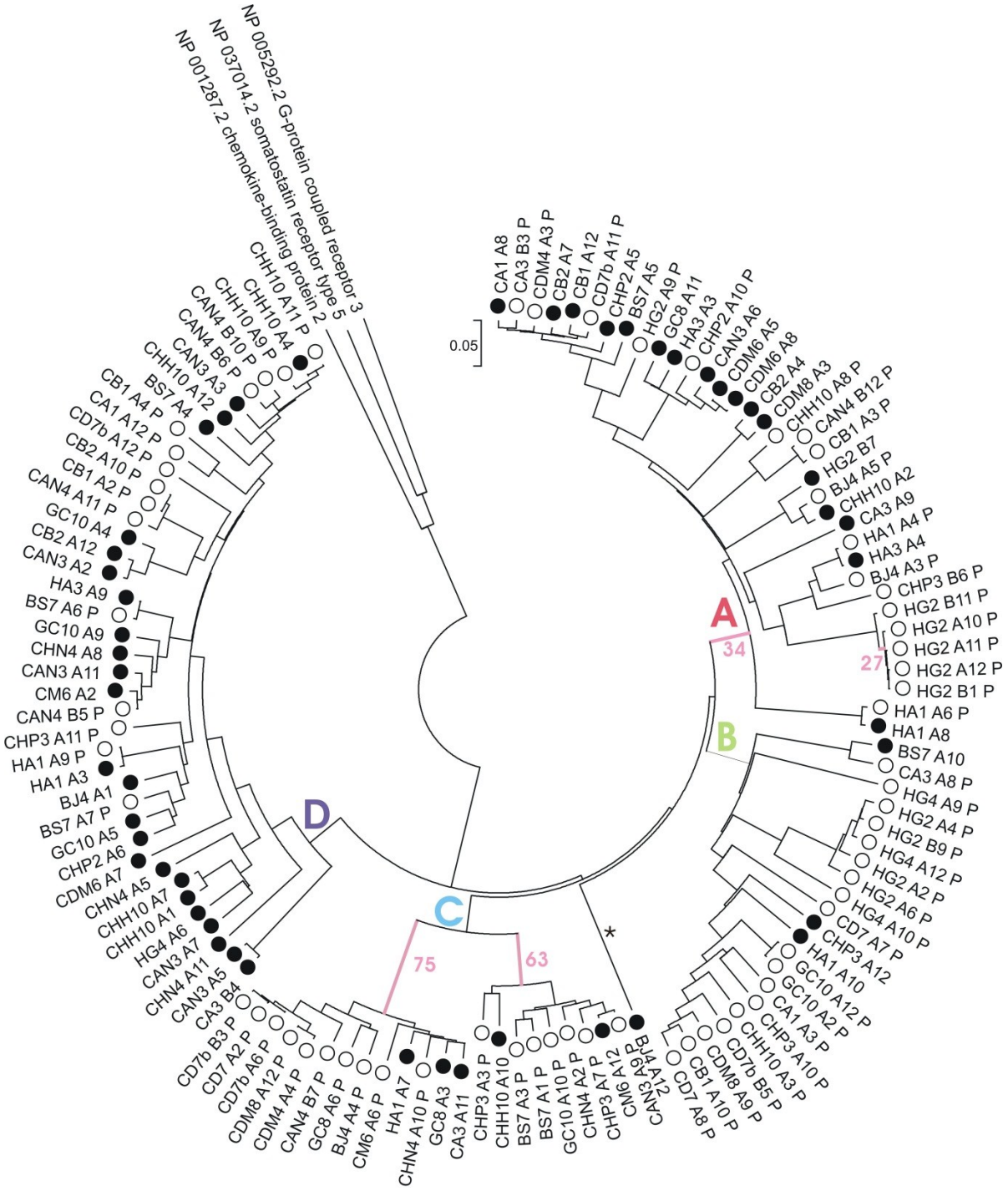
University of Cape Town

Appendix III: Trees

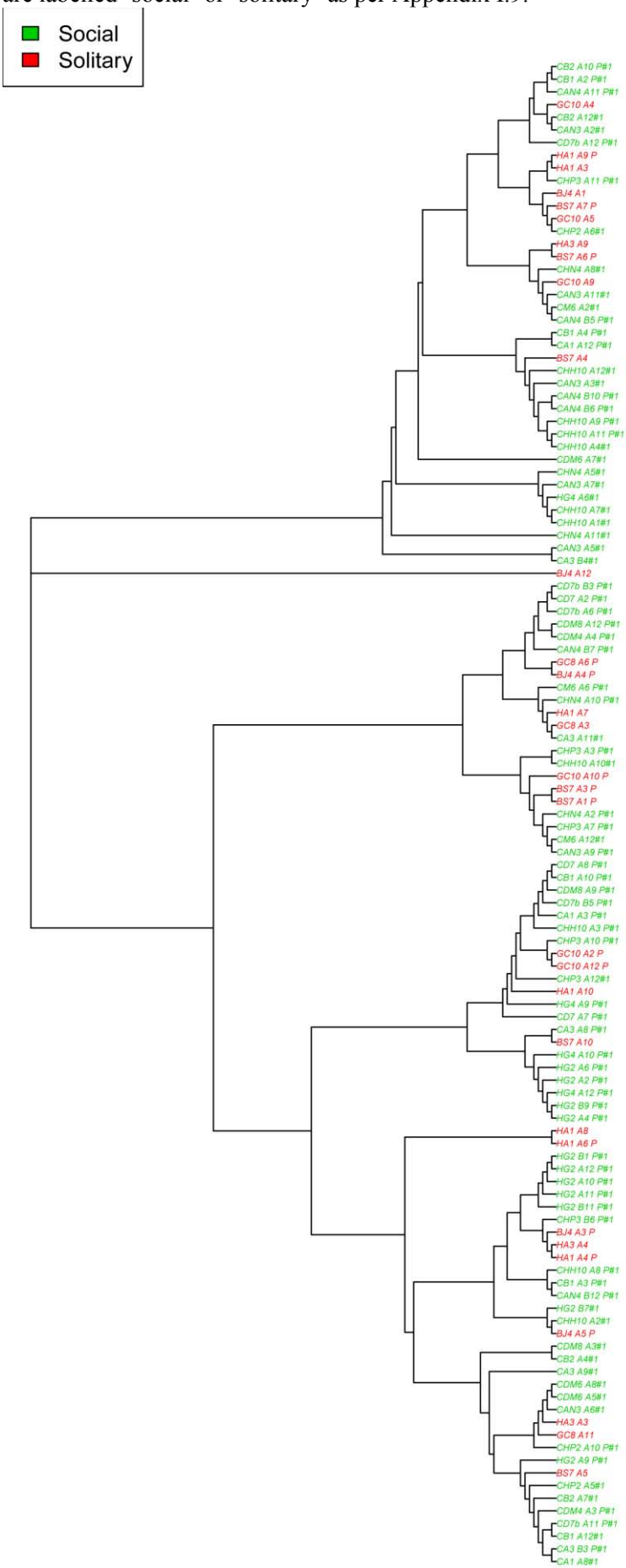
III.1 Bathyergid OR gene tree Maximum-likelihood tree (GTR, 1000 bootstrap) constructed using all 178 unique Bathyergid OR sequences; three rhodopsin-like GPCRs are used as outgroups (accession numbers NP_001287.2, NP_005292.2, NP_037014.2). The four main OR clades are indicated: A (red), B (green), C (blue), D (purple); only one isolated gene (BJ4_A12) falls out of these clades and is labelled with an asterisk. Abbreviations correspond to gene names as per Appendix I.2.



III.2 Bathyergid OR gene tree without allelic variants Maximum-likelihood tree (GTR, 1000 bootstrap) constructed using a single representative sequence for each putative Bathyergid OR gene; three rhodopsin-like GPCRs are used to root the tree (accession numbers NP_001287.2, NP_005292.2, NP_037014.2). The four main OR clades are indicated: A (red), B (green), C (blue), D (purple); only one isolated gene (BJ4_A12) falls out of these clades and is labelled with an asterisk. Abbreviations correspond to gene names as per Appendix I.2. Branches highlighted in pink with the respective branch numbers (also pink) correspond to positively selected branches according to the branch-sites analysis (Yang and Nielsen 2002, Zhang et al. 2005) performed in Chapter 3.



III.3 Partitioned gene-tree Input bathyergid OR gene-tree used to test the Sociality Hypothesis with the methods described by i) Ramm et al. 2008, ii) O'Connor and Mundy 2009, and iii) Mayrose and Otto 2011; terminal branches are labelled 'social' or 'solitary' as per Appendix I.9.



Cape Town

Appendix IV: Scripts

IV.1 Parameter files used to run Ramm et al.'s (2008) method

Null model

```
seqfile = ORnoAGSalignment.phy
treefile = social.nwk
outfile = model.social.H0.out

noisy = 9            * 0,1,2,3,9: how much rubbish on the screen
verbose = 1          * 1: detailed output, 0: concise output
runmode = 0          * 0: user tree;  1: semi-automatic;  2: automatic
                     * 3: StepwiseAddition; (4,5):PerturbationNNI

seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table
clock = 0   * 0: no clock, unrooted tree, 1: clock, rooted tree
model = 2   * models for codons:
              * 0:one, 1:b, 2:2 or more dN/dS ratios for branches

NSsites = 2   * dN/dS among sites. 0:no variation, 1:neutral, 2:positive
icode = 0   * 0:standard genetic code; 1:mammalian mt; 2-10:see below

fix_kappa = 0   * 1: kappa fixed, 0: kappa to be estimated
kappa = 4.54006   * initial or fixed kappa
fix_omega = 1   * 1: omega or omega_1 fixed, 0: estimate
omega = 1   * initial or fixed omega, for codons or codon-transltd AAs

fix_alpha = 1   * 0: estimate gamma shape parameter; 1: fix it at alpha
alpha = .0  * initial or fixed alpha, 0:infinity (constant rate)
Malpha = 0   * different alphas for genes
ncatG = 4   * # of categories in the dG or AdG models of rates

getSE = 0   * 0: don't want them, 1: want S.E.s of estimates
RateAncestor = 0   * (1/0): rates (alpha>0) or ancestral states (alpha=0)

fix_blength = 1  * 0: ignore, -1: random, 1: initial, 2: fixed
method = 0   * 0: simultaneous; 1: one branch at a time

* Specifications for duplicating results for the small data set in table 1
* of Yang (1998 MBE 15:568-573).
* see the tree file lysozyme.trees for specification of node (branch) labels
```

University of Cape Town

Alternative model

```
    seqfile = ORnoAGSalignment.phy
    treefile = social.nwk
    outfile = model.social.H1.out

    noisy = 9            * 0,1,2,3,9: how much rubbish on the screen
    verbose = 1          * 1: detailed output, 0: concise output
    runmode = 0          * 0: user tree;  1: semi-automatic;  2: automatic
                         * 3: StepwiseAddition; (4,5):PerturbationNNI

    seqtype = 1   * 1:codons; 2:AAs; 3:codons-->AAs
    CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table
    clock = 0   * 0: no clock, unrooted tree, 1: clock, rooted tree
    model = 2   * models for codons:
                  * 0:one, 1:b, 2:2 or more dN/dS ratios for branches

    NSsites = 2   * dN/dS among sites. 0:no variation, 1:neutral, 2:positive
    icode = 0   * 0:standard genetic code; 1:mammalian mt; 2-10:see below

    fix_kappa = 0   * 1: kappa fixed, 0: kappa to be estimated
    kappa = 4.54006   * initial or fixed kappa
    fix_omega = 0   * 1: omega or omega_1 fixed, 0: estimate
    omega = 1.1   * initial or fixed omega, for codons or codon-transltd AAs

    fix_alpha = 1   * 0: estimate gamma shape parameter; 1: fix it at alpha
    alpha = .0  * initial or fixed alpha, 0:infinity (constant rate)
    Malpha = 0   * different alphas for genes
    ncatG = 4   * # of categories in the dG or AdG models of rates

    getSE = 0   * 0: don't want them, 1: want S.E.s of estimates
    RateAncestor = 0   * (1/0): rates (alpha>0) or ancestral states (alpha=0)

    fix_blength = 1  * 0: ignore, -1: random, 1: initial, 2: fixed
    method = 0   * 0: simultaneous; 1: one branch at a time

* Specifications for duplicating results for the small data set in table 1
* of Yang (1998 MBE 15:568-573).
* see the tree file lysozyme.trees for specification of node (branch) labels
```

University of Cape Town

IV.2 Parameter files used to run O'Connor and Mundy's (2009) method

Branch length optimisation for phenotype

```
MAXIMUM_ITERATIONS_PER_VARIABLE = 1e26;
LIKELIHOOD_FUNCTION_OUTPUT=1;
DataSet myData = ReadDataFile("alignment.fas");
DataSetFilter myFilter = CreateFilter(myData,1);
HarvestFrequencies(obsFreqs,myFilter,1,1,1);
global a = Random(0,100);
global b = Random(0,100);
global c = Random(0,100);
global d = Random(0,100);
global e = Random(0,100);
global f = Random(0,100);
GTRMatrix = {
{*,t*a,t*b,t*c}
{t*a,*,t*d,t*e}
{t*b,t*d,*,t*f}
{t*c,t*e,t*f,*}
};
Model GTR = (GTRMatrix, obsFreqs);
fscanf("tree.nwk", "String", TREE);
Tree myTree1 = TREE;

LikelihoodFunction like = (myFilter, myTree1, obsFreqs);
Optimize(paramvalues,like);
fprintf(stdout,like);
```

Phenotype rate parameter optimisation

```
MAXIMUM_ITERATIONS_PER_VARIABLE = 1e26;
LIKELIHOOD_FUNCTION_OUTPUT=1;
DataSet myData = ReadDataFile("alignment_phenotype.hyphy");
DataSetFilter myFilter = CreateFilter(myData,1);
HarvestFrequencies(obsFreqs,myFilter,1,1,1);

global p1 = Random(0,100);
Matrix = {
{*,t*p1}
{t*p1,*}
};

Model m = (Matrix, obsFreqs);

fscanf("tree.nwk", "String", TREE);
Tree myTree1 = TREE;

myTree1.Node113.t:=0.00832366;
myTree1.Node131.t:=0.0135423;
myTree1.Node141.t:=0.00205622;
myTree1.Node167.t:=0.00450733;
myTree1.CHH10_A7.t:=0.000513109;
```

University of Cape Town

myTree1.BJ4_A12.t:=0.0186926;
myTree1.Node1.t:=0.00320609;
myTree1.CA3_B4.t:=0.000535016;
myTree1.HG4_A6.t:=0.00620683;
myTree1.CHH10_A1.t:=0.00127601;
myTree1.CAN3_A5.t:=0.000219228;
myTree1.Node158.t:=0.0166474;
myTree1.CHN4_A11.t:=0.0127706;
myTree1.Node145.t:=0.000865442;
myTree1.Node153.t:=0;
myTree1.Node151.t:=0.00100721;
myTree1.Node147.t:=0.00324536;
myTree1.CD7b_B3_P.t:=0.000250653;
myTree1.CD7b_A6_P.t:=0.000250518;
myTree1.CD7_A2_P.t:=0.000250319;
myTree1.Node166.t:=0.00103339;
myTree1.Node181.t:=0.000264646;
myTree1.CAN4_B6_P.t:=0.000765665;
myTree1.CAN4_B10_P.t:=0.000757384;
myTree1.CHH10_A9_P.t:=0.000505776;
myTree1.CHH10_A11_P.t:=0;
myTree1.Node182.t:=0.000507958;
myTree1.Node186.t:=0.00153142;
myTree1.CHN4_A5.t:=0.00736816;
myTree1.Node165.t:=0.0032995;
myTree1.CAN3_A7.t:=0.00548473;
myTree1.Node164.t:=0.00433004;
myTree1.CHH10_A4.t:=0;
myTree1.CDM6_A7.t:=0.0112384;
myTree1.Node148.t:=0.000250556;
myTree1.CHH10_A10.t:=0.00549282;
myTree1.Node115.t:=0.00277585;
myTree1.GC10_A10_P.t:=0.00251972;
myTree1.CHP3_A3_P.t:=0.00178408;
myTree1.CA3_A11.t:=0.0020163;
myTree1.Node114.t:=0.0111347;
myTree1.Node128.t:=0.00175277;
myTree1.Node116.t:=0;
myTree1.BS7_A1_P.t:=0.00125993;
myTree1.BS7_A3_P.t:=0.00256663;
myTree1.Node124.t:=0.00124771;
myTree1.Node117.t:=0.000262073;
myTree1.Node118.t:=0.000506046;
myTree1.CHN4_A2_P.t:=0.00152132;
myTree1.GC8_A3.t:=0.0012246;
myTree1.Node142.t:=0.000470719;
myTree1.GC8_A6_P.t:=0.00270684;
myTree1.BJ4_A4_P.t:=0.00219152;
myTree1.CAN4_B7_P.t:=0.00219348;
myTree1.CDM8_A12_P.t:=0;
myTree1.CDM4_A4_P.t:=0.000753589;
myTree1.Node132.t:=0.000245948;
myTree1.Node134.t:=0;

University of Cape Town

myTree1.HA1_A7.t:=0.000850229;
myTree1.Node135.t:=0.000523328;
myTree1.CHN4_A10_P.t:=0.00294701;
myTree1.CM6_A6_P.t:=0.00466808;
myTree1.Node133.t:=0.000275251;
myTree1.Node180.t:=0.00022233;
myTree1.CAN3_A2.t:=0.000764794;
myTree1.CD7b_A12_P.t:=0.00697013;
myTree1.Node210.t:=0.00159005;
myTree1.CB2_A12.t:=0.000767318;
myTree1.Node226.t:=0.00391231;
myTree1.GC10_A4.t:=0.00102171;
myTree1.Node227.t:=0.00157008;
myTree1.Node218.t:=0.00184429;
myTree1.CHP3_A11_P.t:=0.00570809;
myTree1.Node211.t:=0.00112462;
myTree1.BJ4_A1.t:=0.00521648;
myTree1.HA1_A3.t:=0.00025275;
myTree1.Node220.t:=0.00925611;
myTree1.HA1_A9_P.t:=0;
myTree1.CAN4_A11_P.t:=0.00103152;
myTree1.Node173.t:=0.00423551;
myTree1.Node175.t:=0.00286773;
myTree1.Node195.t:=0.000702924;
myTree1.Node163.t:=0.00250534;
myTree1.Node157.t:=0.00863618;
myTree1.Node161.t:=0.0121764;
myTree1.Node209.t:=0.000586838;
myTree1.Node233.t:=0.000504384;
myTree1.CB2_A10_P.t:=0.000254366;
myTree1.CB1_A2_P.t:=0;
myTree1.Node231.t:=0.00590885;
myTree1.Node223.t:=0.00207671;
myTree1.Node225.t:=0.00135662;
myTree1.Node212.t:=0.00273539;
myTree1.Node176.t:=0.00162311;
myTree1.Node192.t:=0.00681386;
myTree1.CB1_A4_P.t:=0.00445268;
myTree1.CAN4_B5_P.t:=0;
myTree1.Node200.t:=0.000410666;
myTree1.CM6_A2.t:=0.00101263;
myTree1.CA1_A12_P.t:=0.00195191;
myTree1.CHH10_A12.t:=0.00232827;
myTree1.Node179.t:=0.000802216;
myTree1.CAN3_A3.t:=0.00232951;
myTree1.Node178.t:=0.00207634;
myTree1.Node177.t:=0.00269063;
myTree1.BS7_A4.t:=0.00563011;
myTree1.CAN3_A11.t:=0.00138153;
myTree1.CHP2_A6.t:=0.00301277;
myTree1.Node196.t:=0.00334811;
myTree1.Node206.t:=0.00310539;
myTree1.GC10_A5.t:=0.00271463;

University of Cape Town

myTree1.BS7_A7_P.t:=0.00295726;
myTree1.Node213.t:=0.00131694;
myTree1.HA3_A9.t:=0;
myTree1.Node198.t:=0.00128282;
myTree1.GC10_A9.t:=0.00404663;
myTree1.Node199.t:=0.00116455;
myTree1.CHN4_A8.t:=0.00297054;
myTree1.BS7_A6_P.t:=0.000504451;
myTree1.Node197.t:=0.00205149;
myTree1.CHP3_A7_P.t:=0.0020356;
myTree1.HG2_B7.t:=0.00273378;
myTree1.Node42.t:=0.00226424;
myTree1.CHH10_A2.t:=0.00119187;
myTree1.Node41.t:=0.0112604;
myTree1.Node48.t:=0.00602513;
myTree1.CB1_A3_P.t:=0.000781089;
myTree1.CAN4_B12_P.t:=0.000236488;
myTree1.BJ4_A5_P.t:=0.00083322;
myTree1.CB2_A4.t:=0.00148122;
myTree1.Node6.t:=0.000923487;
myTree1.CA3_A9.t:=0.014179;
myTree1.CDM8_A3.t:=0.00107823;
myTree1.Node5.t:=0.00095855;
myTree1.Node37.t:=0.0115932;
myTree1.CHH10_A8_P.t:=0.000610863;
myTree1.HG2_A11_P.t:=0;
myTree1.HG2_B11_P.t:=0.000222962;
myTree1.Node53.t:=0.00592811;
myTree1.HG2_A10_P.t:=0.000249577;
myTree1.HG2_B1_P.t:=0;
myTree1.HG2_A12_P.t:=0;
myTree1.CHP3_B6_P.t:=0.00636273;
myTree1.HA3_A4.t:=0;
myTree1.HA1_A4_P.t:=0.000248571;
myTree1.Node47.t:=0.0109742;
myTree1.Node55.t:=0.00391241;
myTree1.Node54.t:=0.0010105;
myTree1.BJ4_A3_P.t:=0.00137033;
myTree1.Node7.t:=0.00212069;
myTree1.CB2_A7.t:=0.000751243;
myTree1.Node12.t:=0;
myTree1.CDM4_A3_P.t:=0.00050051;
myTree1.Node11.t:=0.000509885;
myTree1.BS7_A5.t:=0.00323298;
myTree1.Node10.t:=0.00117747;
myTree1.CHP2_A5.t:=0.00150956;
myTree1.Node13.t:=0;
myTree1.Node14.t:=0.000503596;
myTree1.CA3_B3_P.t:=0.000754232;
myTree1.CA1_A8.t:=0.00125559;
myTree1.CB1_A12.t:=0.000501086;
myTree1.Node17.t:=0.000502439;
myTree1.CD7b_A11_P.t:=0.00024925;

University of Cape Town

myTree1.Node9.t:=0.0039536;
myTree1.Node31.t:=0.000418098;
myTree1.Node33.t:=0.000515897;
myTree1.CDM6_A8.t:=0.000250097;
myTree1.Node29.t:=0.000125062;
myTree1.Node25.t:=0.00792167;
myTree1.Node27.t:=0.00041945;
myTree1.CDM6_A5.t:=0;
myTree1.CHP2_A10_P.t:=0.0026468;
myTree1.Node8.t:=0.00335462;
myTree1.HG2_A9_P.t:=0.00421091;
myTree1.GC8_A11.t:=0.0022704;
myTree1.CAN3_A6.t:=0.00150208;
myTree1.HA3_A3.t:=0.00276997;
myTree1.Node66.t:=0;
myTree1.CD7b_B5_P.t:=0.00281446;
myTree1.CA1_A3_P.t:=0.00679279;
myTree1.CHH10_A3_P.t:=0.00496893;
myTree1.CDM8_A9_P.t:=0.00165468;
myTree1.Node110.t:=0.00101109;
myTree1.CD7_A8_P.t:=0.00164329;
myTree1.CB1_A10_P.t:=0.000643581;
myTree1.Node97.t:=0.00174671;
myTree1.GC10_A12_P.t:=0.000251826;
myTree1.CHP3_A12.t:=0.00802465;
myTree1.HA1_A10.t:=0.00862744;
myTree1.GC10_A2_P.t:=0.000253429;
myTree1.CHP3_A10_P.t:=0.00644101;
myTree1.Node98.t:=0.00437395;
myTree1.Node108.t:=0.00167792;
myTree1.Node2.t:=0.00390231;
myTree1.Node72.t:=0.0135382;
myTree1.Node88.t:=0.00288093;
myTree1.CAN3_A9_P.t:=0.000503784;
myTree1.Node119.t:=0.00050282;
myTree1.CM6_A12.t:=0.000503392;
myTree1.Node90.t:=0.000925552;
myTree1.Node102.t:=0.0021355;
myTree1.Node104.t:=0.00102634;
myTree1.Node106.t:=0;
myTree1.Node96.t:=0.00281281;
myTree1.Node92.t:=0.00576115;
myTree1.Node94.t:=0.000663541;
myTree1.HG4_A9_P.t:=0.0163667;
myTree1.Node69.t:=0.0151686;
myTree1.HA1_A8.t:=0.00027018;
myTree1.HA1_A6_P.t:=0.000740252;
myTree1.Node3.t:=0.0116403;
myTree1.HG2_B9_P.t:=0.000355084;
myTree1.HG2_A4_P.t:=0.000907729;
myTree1.Node4.t:=0.000346362;
myTree1.Node60.t:=0.0115746;
myTree1.Node62.t:=0.00103971;

University of Cape Town

```
myTree1.Node64.t:=0.00024987;
myTree1.Node52.t:=0.00152817;
myTree1.Node40.t:=0.00218827;
myTree1.Node46.t:=0.00146772;
myTree1.Node78.t:=0.000222681;
myTree1.CA3_A8_P.t:=0.00327665;
myTree1.BS7_A10.t:=0.00160172;
myTree1.Node74.t:=0.0073392;
myTree1.Node85.t:=0.0143198;
myTree1.CD7_A7_P.t:=0.0155596;
myTree1.Node73.t:=0.00390712;
myTree1.HG4_A10_P.t:=0.00739958;
myTree1.HG2_A2_P.t:=0.00175202;
myTree1.Node77.t:=0.00108641;
myTree1.HG4_A12_P.t:=0.00093638;
myTree1.Node76.t:=0.00113484;
myTree1.Node75.t:=0.00548586;
myTree1.HG2_A6_P.t:=0.00178151;


LikelihoodFunction like = (myFilter, myTree1);
Optimize(paramvalues,like);
fprintf(stdout,like);

# Null model analysis

MAXIMUM_ITERATIONS_PER_VARIABLE = 1e26;
LIKELIHOOD_FUNCTION_OUTPUT=1;
DataSet myData = ReadDataFile("alignment_solitary.hyphy");
DataSetFilter myFilter = CreateFilter(myData,1);
HarvestFrequencies(obsFreqs,myFilter,1,1,1);
fprintf(stdout,obsFreqs,"\n");
global W0 = Random(0,100);
global W1 = Random(0,100);
global a = Random(0,100);
global b = Random(0,100);
global c = Random(0,100);
global d = Random(0,100);
global e = Random(0,100);
global f = Random(0,100);
global p1 := 380.351;


IndependentRateMatrix = {
{*,t*p1,tt*a,0,tt*b,0,tt*c,0}
{t*p1,*,0,tt*a,0,tt*b,0,tt*c}
{tt*a,0,*,t*p1,tt*d,0,tt*e,0}
{0,tt*a,t*p1,*,0,tt*d,0,tt*e}
{tt*b,0,tt*d,0,*,t*p1,tt*f,0}
{0,tt*b,0,tt*d,t*p1,*,0,tt*f}
{tt*c,0,tt*e,0,tt*f,0,*,t*p1}
{0,tt*c,0,tt*e,0,tt*f,t*p1,*}
};
```

University of Cape Town

```
Model Ind = (IndependentRateMatrix, obsFreqs);
fscanf("tree.nwk", "String", TREE);
Tree myTree1 = TREE;

myTree1.Node113.t:=0.00832366;
myTree1.Node131.t:=0.0135423;
myTree1.Node141.t:=0.00205622;
myTree1.Node167.t:=0.00450733;
myTree1.CHH10_A7.t:=0.000513109;
myTree1.BJ4_A12.t:=0.0186926;
myTree1.Node1.t:=0.00320609;
myTree1.CA3_B4.t:=0.000535016;
myTree1.HG4_A6.t:=0.00620683;
myTree1.CHH10_A1.t:=0.00127601;
myTree1.CAN3_A5.t:=0.000219228;
myTree1.Node158.t:=0.0166474;
myTree1.CHN4_A11.t:=0.0127706;
myTree1.Node145.t:=0.000865442;
myTree1.Node153.t:=0;
myTree1.Node151.t:=0.00100721;
myTree1.Node147.t:=0.00324536;
myTree1.CD7b_B3_P.t:=0.000250653;
myTree1.CD7b_A6_P.t:=0.000250518;
myTree1.CD7_A2_P.t:=0.000250319;
myTree1.Node166.t:=0.00103339;
myTree1.Node181.t:=0.000264646;
myTree1.CAN4_B6_P.t:=0.000765665;
myTree1.CAN4_B10_P.t:=0.000757384;
myTree1.CHH10_A9_P.t:=0.000505776;
myTree1.CHH10_A11_P.t:=0;
myTree1.Node182.t:=0.000507958;
myTree1.Node186.t:=0.00153142;
myTree1.CHN4_A5.t:=0.00736816;
myTree1.Node165.t:=0.0032995;
myTree1.CAN3_A7.t:=0.00548473;
myTree1.Node164.t:=0.00433004;
myTree1.CHH10_A4.t:=0;
myTree1.CDM6_A7.t:=0.0112384;
myTree1.Node148.t:=0.000250556;
myTree1.CHH10_A10.t:=0.00549282;
myTree1.Node115.t:=0.00277585;
myTree1.GC10_A10_P.t:=0.00251972;
myTree1.CHP3_A3_P.t:=0.00178408;
myTree1.CA3_A11.t:=0.0020163;
myTree1.Node114.t:=0.0111347;
myTree1.Node128.t:=0.00175277;
myTree1.Node116.t:=0;
myTree1.BS7_A1_P.t:=0.00125993;
myTree1.BS7_A3_P.t:=0.00256663;
myTree1.Node124.t:=0.00124771;
myTree1.Node117.t:=0.000262073;
myTree1.Node118.t:=0.000506046;
```

University of Cape Town

myTree1.CHN4_A2_P.t:=0.00152132;
myTree1.GC8_A3.t:=0.0012246;
myTree1.Node142.t:=0.000470719;
myTree1.GC8_A6_P.t:=0.00270684;
myTree1.BJ4_A4_P.t:=0.00219152;
myTree1.CAN4_B7_P.t:=0.00219348;
myTree1.CDM8_A12_P.t:=0;
myTree1.CDM4_A4_P.t:=0.000753589;
myTree1.Node132.t:=0.000245948;
myTree1.Node134.t:=0;
myTree1.HA1_A7.t:=0.000850229;
myTree1.Node135.t:=0.000523328;
myTree1.CHN4_A10_P.t:=0.00294701;
myTree1.CM6_A6_P.t:=0.00466808;
myTree1.Node133.t:=0.000275251;
myTree1.Node180.t:=0.00022233;
myTree1.CAN3_A2.t:=0.000764794;
myTree1.CD7b_A12_P.t:=0.00697013;
myTree1.Node210.t:=0.00159005;
myTree1.CB2_A12.t:=0.000767318;
myTree1.Node226.t:=0.00391231;
myTree1.GC10_A4.t:=0.00102171;
myTree1.Node227.t:=0.00157008;
myTree1.Node218.t:=0.00184429;
myTree1.CHP3_A11_P.t:=0.00570809;
myTree1.Node211.t:=0.00112462;
myTree1.BJ4_A1.t:=0.00521648;
myTree1.HA1_A3.t:=0.00025275;
myTree1.Node220.t:=0.00925611;
myTree1.HA1_A9_P.t:=0;
myTree1.CAN4_A11_P.t:=0.00103152;
myTree1.Node173.t:=0.00423551;
myTree1.Node175.t:=0.00286773;
myTree1.Node195.t:=0.000702924;
myTree1.Node163.t:=0.00250534;
myTree1.Node157.t:=0.00863618;
myTree1.Node161.t:=0.0121764;
myTree1.Node209.t:=0.000586838;
myTree1.Node233.t:=0.000504384;
myTree1.CB2_A10_P.t:=0.000254366;
myTree1.CB1_A2_P.t:=0;
myTree1.Node231.t:=0.00590885;
myTree1.Node223.t:=0.00207671;
myTree1.Node225.t:=0.00135662;
myTree1.Node212.t:=0.00273539;
myTree1.Node176.t:=0.00162311;
myTree1.Node192.t:=0.00681386;
myTree1.CB1_A4_P.t:=0.00445268;
myTree1.CAN4_B5_P.t:=0;
myTree1.Node200.t:=0.000410666;
myTree1.CM6_A2.t:=0.00101263;
myTree1.CA1_A12_P.t:=0.00195191;
myTree1.CHH10_A12.t:=0.00232827;

University of Cape Town

myTree1.Node179.t:=0.000802216;
myTree1.CAN3_A3.t:=0.00232951;
myTree1.Node178.t:=0.00207634;
myTree1.Node177.t:=0.00269063;
myTree1.BS7_A4.t:=0.00563011;
myTree1.CAN3_A11.t:=0.00138153;
myTree1.CHP2_A6.t:=0.00301277;
myTree1.Node196.t:=0.00334811;
myTree1.Node206.t:=0.00310539;
myTree1.GC10_A5.t:=0.00271463;
myTree1.BS7_A7_P.t:=0.00295726;
myTree1.Node213.t:=0.00131694;
myTree1.HA3_A9.t:=0;
myTree1.Node198.t:=0.00128282;
myTree1.GC10_A9.t:=0.00404663;
myTree1.Node199.t:=0.00116455;
myTree1.CHN4_A8.t:=0.00297054;
myTree1.BS7_A6_P.t:=0.000504451;
myTree1.Node197.t:=0.00205149;
myTree1.CHP3_A7_P.t:=0.0020356;
myTree1.HG2_B7.t:=0.00273378;
myTree1.Node42.t:=0.00226424;
myTree1.CHH10_A2.t:=0.00119187;
myTree1.Node41.t:=0.0112604;
myTree1.Node48.t:=0.00602513;
myTree1.CB1_A3_P.t:=0.000781089;
myTree1.CAN4_B12_P.t:=0.000236488;
myTree1.BJ4_A5_P.t:=0.00083322;
myTree1.CB2_A4.t:=0.00148122;
myTree1.Node6.t:=0.000923487;
myTree1.CA3_A9.t:=0.014179;
myTree1.CDM8_A3.t:=0.00107823;
myTree1.Node5.t:=0.00095855;
myTree1.Node37.t:=0.0115932;
myTree1.CHH10_A8_P.t:=0.000610863;
myTree1.HG2_A11_P.t:=0;
myTree1.HG2_B11_P.t:=0.000222962;
myTree1.Node53.t:=0.00592811;
myTree1.HG2_A10_P.t:=0.000249577;
myTree1.HG2_B1_P.t:=0;
myTree1.HG2_A12_P.t:=0;
myTree1.CHP3_B6_P.t:=0.00636273;
myTree1.HA3_A4.t:=0;
myTree1.HA1_A4_P.t:=0.000248571;
myTree1.Node47.t:=0.0109742;
myTree1.Node55.t:=0.00391241;
myTree1.Node54.t:=0.0010105;
myTree1.BJ4_A3_P.t:=0.00137033;
myTree1.Node7.t:=0.00212069;
myTree1.CB2_A7.t:=0.000751243;
myTree1.Node12.t:=0;
myTree1.CDM4_A3_P.t:=0.00050051;
myTree1.Node11.t:=0.000509885;

University of Cape Town

myTree1.BS7_A5.t:=0.00323298;
myTree1.Node10.t:=0.00117747;
myTree1.CHP2_A5.t:=0.00150956;
myTree1.Node13.t:=0;
myTree1.Node14.t:=0.000503596;
myTree1.CA3_B3_P.t:=0.000754232;
myTree1.CA1_A8.t:=0.00125559;
myTree1.CB1_A12.t:=0.000501086;
myTree1.Node17.t:=0.000502439;
myTree1.CD7b_A11_P.t:=0.00024925;
myTree1.Node9.t:=0.0039536;
myTree1.Node31.t:=0.000418098;
myTree1.Node33.t:=0.000515897;
myTree1.CDM6_A8.t:=0.000250097;
myTree1.Node29.t:=0.000125062;
myTree1.Node25.t:=0.00792167;
myTree1.Node27.t:=0.00041945;
myTree1.CDM6_A5.t:=0;
myTree1.CHP2_A10_P.t:=0.0026468;
myTree1.Node8.t:=0.00335462;
myTree1.HG2_A9_P.t:=0.00421091;
myTree1.GC8_A11.t:=0.0022704;
myTree1.CAN3_A6.t:=0.00150208;
myTree1.HA3_A3.t:=0.00276997;
myTree1.Node66.t:=0;
myTree1.CD7b_B5_P.t:=0.00281446;
myTree1.CA1_A3_P.t:=0.00679279;
myTree1.CHH10_A3_P.t:=0.00496893;
myTree1.CDM8_A9_P.t:=0.00165468;
myTree1.Node110.t:=0.00101109;
myTree1.CD7_A8_P.t:=0.00164329;
myTree1.CB1_A10_P.t:=0.000643581;
myTree1.Node97.t:=0.00174671;
myTree1.GC10_A12_P.t:=0.000251826;
myTree1.CHP3_A12.t:=0.00802465;
myTree1.HA1_A10.t:=0.00862744;
myTree1.GC10_A2_P.t:=0.000253429;
myTree1.CHP3_A10_P.t:=0.00644101;
myTree1.Node98.t:=0.00437395;
myTree1.Node108.t:=0.00167792;
myTree1.Node2.t:=0.00390231;
myTree1.Node72.t:=0.0135382;
myTree1.Node88.t:=0.00288093;
myTree1.CAN3_A9_P.t:=0.000503784;
myTree1.Node119.t:=0.00050282;
myTree1.CM6_A12.t:=0.000503392;
myTree1.Node90.t:=0.000925552;
myTree1.Node102.t:=0.0021355;
myTree1.Node104.t:=0.00102634;
myTree1.Node106.t:=0;
myTree1.Node96.t:=0.00281281;
myTree1.Node92.t:=0.00576115;
myTree1.Node94.t:=0.000663541;

University of Cape Town

```
myTree1.HG4_A9_P.t:=0.0163667;
myTree1.Node69.t:=0.0151686;
myTree1.HA1_A8.t:=0.00027018;
myTree1.HA1_A6_P.t:=0.000740252;
myTree1.Node3.t:=0.0116403;
myTree1.HG2_B9_P.t:=0.000355084;
myTree1.HG2_A4_P.t:=0.000907729;
myTree1.Node4.t:=0.000346362;
myTree1.Node60.t:=0.0115746;
myTree1.Node62.t:=0.00103971;
myTree1.Node64.t:=0.00024987;
myTree1.Node52.t:=0.00152817;
myTree1.Node40.t:=0.00218827;
myTree1.Node46.t:=0.00146772;
myTree1.Node78.t:=0.000222681;
myTree1.CA3_A8_P.t:=0.00327665;
myTree1.BS7_A10.t:=0.00160172;
myTree1.Node74.t:=0.0073392;
myTree1.Node85.t:=0.0143198;
myTree1.CD7_A7_P.t:=0.0155596;
myTree1.Node73.t:=0.00390712;
myTree1.HG4_A10_P.t:=0.00739958;
myTree1.HG2_A2_P.t:=0.00175202;
myTree1.Node77.t:=0.00108641;
myTree1.HG4_A12_P.t:=0.00093638;
myTree1.Node76.t:=0.00113484;
myTree1.Node75.t:=0.00548586;
myTree1.HG2_A6_P.t:=0.00178151;


W0 := W1;
DependentRateMatrix = {
{*,t*p1,W0*tt*a,0,W0*tt*b,0,W0*tt*c,0}
{t*p1,*,0,W1*tt*a,0,W1*tt*b,0,W1*tt*c}
{W0*tt*a,0,*,t*p1,W0*tt*d,0,W0*tt*e,0}
{0,W1*tt*a,t*p1,*,0,W1*tt*d,0,W1*tt*e}
{W0*tt*b,0,W0*tt*d,0,*,t*p1,W0*tt*f,0}
{0,W1*tt*b,0,W1*tt*d,t*p1,*,0,W1*tt*f}
{W0*tt*c,0,W0*tt*e,0,W0*tt*f,0,*,t*p1}
{0,W1*tt*c,0,W1*tt*e,0,W1*tt*f,t*p1,*}
};
Model Dep = (DependentRateMatrix, obsFreqs);
Tree myTree2 = TREE;
global P_1 = Random(0,1);
P_1 :< 1;
bn = BranchName(myTree1,-1);
for(k=0; k< Columns(bn) -1; k=k+1){
        outString = "myTree2."+bn[k]+".t := myTree1."+bn[k]+".t";
        ExecuteCommands(outString);
        outString = "myTree2."+bn[k]+".tt := myTree1."+bn[k]+".tt";
        ExecuteCommands(outString);
}
```

University of Cape Town

```
LikelihoodFunction theLikFun = (myFilter, myTree1,myFilter, myTree2,
"Log((1-P_1)*SITE_LIKELIHOOD[0]+P_1*SITE_LIKELIHOOD[1])");
Optimize(results,theLikFun);
fprintf(stdout,theLikFun);
```

Alternative model analysis

```
MAXIMUM_ITERATIONS_PER_VARIABLE = 1e26;
LIKELIHOOD_FUNCTION_OUTPUT=1;
DataSet myData = ReadDataFile("alignment_solitary.hyphy");
DataSetFilter myFilter = CreateFilter(myData,1);
HarvestFrequencies(obsFreqs,myFilter,1,1,1);
fprintf(stdout,obsFreqs,"\n");
global W0 = Random(0,100);
global W1 = Random(0,100);
global a = Random(0,100);
global b = Random(0,100);
global c = Random(0,100);
global d = Random(0,100);
global e = Random(0,100);
global f = Random(0,100);
global p1 := 380.351;

IndependentRateMatrix = {
{*,t*p1,tt*a,0,tt*b,0,tt*c,0}
{t*p1,*,0,tt*a,0,tt*b,0,tt*c}
{tt*a,0,*,t*p1,tt*d,0,tt*e,0}
{0,tt*a,t*p1,*,0,tt*d,0,tt*e}
{tt*b,0,tt*d,0,*,t*p1,tt*f,0}
{0,tt*b,0,tt*d,t*p1,*,0,tt*f}
{tt*c,0,tt*e,0,tt*f,0,*,t*p1}
{0,tt*c,0,tt*e,0,tt*f,t*p1,*}
};
Model Ind = (IndependentRateMatrix, obsFreqs);

fscanf("tree.nwk", "String", TREE);
Tree myTree1 = TREE;

myTree1.Node113.t:=0.00832366;
myTree1.Node131.t:=0.0135423;
myTree1.Node141.t:=0.00205622;
myTree1.Node167.t:=0.00450733;
myTree1.CHH10_A7.t:=0.000513109;
myTree1.BJ4_A12.t:=0.0186926;
myTree1.Node1.t:=0.00320609;
myTree1.CA3_B4.t:=0.000535016;
myTree1.HG4_A6.t:=0.00620683;
myTree1.CHH10_A1.t:=0.00127601;
myTree1.CAN3_A5.t:=0.000219228;
myTree1.Node158.t:=0.0166474;
myTree1.CHN4_A11.t:=0.0127706;
myTree1.Node145.t:=0.000865442;
myTree1.Node153.t:=0;
```

University of Cape Town

myTree1.Node151.t:=0.00100721;
myTree1.Node147.t:=0.00324536;
myTree1.CD7b_B3_P.t:=0.000250653;
myTree1.CD7b_A6_P.t:=0.000250518;
myTree1.CD7_A2_P.t:=0.000250319;
myTree1.Node166.t:=0.00103339;
myTree1.Node181.t:=0.000264646;
myTree1.CAN4_B6_P.t:=0.000765665;
myTree1.CAN4_B10_P.t:=0.000757384;
myTree1.CHH10_A9_P.t:=0.000505776;
myTree1.CHH10_A11_P.t:=0;
myTree1.Node182.t:=0.000507958;
myTree1.Node186.t:=0.00153142;
myTree1.CHN4_A5.t:=0.00736816;
myTree1.Node165.t:=0.0032995;
myTree1.CAN3_A7.t:=0.00548473;
myTree1.Node164.t:=0.00433004;
myTree1.CHH10_A4.t:=0;
myTree1.CDM6_A7.t:=0.0112384;
myTree1.Node148.t:=0.000250556;
myTree1.CHH10_A10.t:=0.00549282;
myTree1.Node115.t:=0.00277585;
myTree1.GC10_A10_P.t:=0.00251972;
myTree1.CHP3_A3_P.t:=0.00178408;
myTree1.CA3_A11.t:=0.0020163;
myTree1.Node114.t:=0.0111347;
myTree1.Node128.t:=0.00175277;
myTree1.Node116.t:=0;
myTree1.BS7_A1_P.t:=0.00125993;
myTree1.BS7_A3_P.t:=0.00256663;
myTree1.Node124.t:=0.00124771;
myTree1.Node117.t:=0.000262073;
myTree1.Node118.t:=0.000506046;
myTree1.CHN4_A2_P.t:=0.00152132;
myTree1.GC8_A3.t:=0.0012246;
myTree1.Node142.t:=0.000470719;
myTree1.GC8_A6_P.t:=0.00270684;
myTree1.BJ4_A4_P.t:=0.00219152;
myTree1.CAN4_B7_P.t:=0.00219348;
myTree1.CDM8_A12_P.t:=0;
myTree1.CDM4_A4_P.t:=0.000753589;
myTree1.Node132.t:=0.000245948;
myTree1.Node134.t:=0;
myTree1.HA1_A7.t:=0.000850229;
myTree1.Node135.t:=0.000523328;
myTree1.CHN4_A10_P.t:=0.00294701;
myTree1.CM6_A6_P.t:=0.00466808;
myTree1.Node133.t:=0.000275251;
myTree1.Node180.t:=0.00022233;
myTree1.CAN3_A2.t:=0.000764794;
myTree1.CD7b_A12_P.t:=0.00697013;
myTree1.Node210.t:=0.00159005;
myTree1.CB2_A12.t:=0.000767318;

University of Cape Town

myTree1.Node226.t:=0.00391231;
myTree1.GC10_A4.t:=0.00102171;
myTree1.Node227.t:=0.00157008;
myTree1.Node218.t:=0.00184429;
myTree1.CHP3_A11_P.t:=0.00570809;
myTree1.Node211.t:=0.00112462;
myTree1.BJ4_A1.t:=0.00521648;
myTree1.HA1_A3.t:=0.00025275;
myTree1.Node220.t:=0.00925611;
myTree1.HA1_A9_P.t:=0;
myTree1.CAN4_A11_P.t:=0.00103152;
myTree1.Node173.t:=0.00423551;
myTree1.Node175.t:=0.00286773;
myTree1.Node195.t:=0.000702924;
myTree1.Node163.t:=0.00250534;
myTree1.Node157.t:=0.00863618;
myTree1.Node161.t:=0.0121764;
myTree1.Node209.t:=0.000586838;
myTree1.Node233.t:=0.000504384;
myTree1.CB2_A10_P.t:=0.000254366;
myTree1.CB1_A2_P.t:=0;
myTree1.Node231.t:=0.00590885;
myTree1.Node223.t:=0.00207671;
myTree1.Node225.t:=0.00135662;
myTree1.Node212.t:=0.00273539;
myTree1.Node176.t:=0.00162311;
myTree1.Node192.t:=0.00681386;
myTree1.CB1_A4_P.t:=0.00445268;
myTree1.CAN4_B5_P.t:=0;
myTree1.Node200.t:=0.000410666;
myTree1.CM6_A2.t:=0.00101263;
myTree1.CA1_A12_P.t:=0.00195191;
myTree1.CHH10_A12.t:=0.00232827;
myTree1.Node179.t:=0.000802216;
myTree1.CAN3_A3.t:=0.00232951;
myTree1.Node178.t:=0.00207634;
myTree1.Node177.t:=0.00269063;
myTree1.BS7_A4.t:=0.00563011;
myTree1.CAN3_A11.t:=0.00138153;
myTree1.CHP2_A6.t:=0.00301277;
myTree1.Node196.t:=0.00334811;
myTree1.Node206.t:=0.00310539;
myTree1.GC10_A5.t:=0.00271463;
myTree1.BS7_A7_P.t:=0.00295726;
myTree1.Node213.t:=0.00131694;
myTree1.HA3_A9.t:=0;
myTree1.Node198.t:=0.00128282;
myTree1.GC10_A9.t:=0.00404663;
myTree1.Node199.t:=0.00116455;
myTree1.CHN4_A8.t:=0.00297054;
myTree1.BS7_A6_P.t:=0.000504451;
myTree1.Node197.t:=0.00205149;
myTree1.CHP3_A7_P.t:=0.0020356;

University of Cape Town

myTree1.HG2_B7.t:=0.00273378;
myTree1.Node42.t:=0.00226424;
myTree1.CHH10_A2.t:=0.00119187;
myTree1.Node41.t:=0.0112604;
myTree1.Node48.t:=0.00602513;
myTree1.CB1_A3_P.t:=0.000781089;
myTree1.CAN4_B12_P.t:=0.000236488;
myTree1.BJ4_A5_P.t:=0.00083322;
myTree1.CB2_A4.t:=0.00148122;
myTree1.Node6.t:=0.000923487;
myTree1.CA3_A9.t:=0.014179;
myTree1.CDM8_A3.t:=0.00107823;
myTree1.Node5.t:=0.00095855;
myTree1.Node37.t:=0.0115932;
myTree1.CHH10_A8_P.t:=0.000610863;
myTree1.HG2_A11_P.t:=0;
myTree1.HG2_B11_P.t:=0.000222962;
myTree1.Node53.t:=0.00592811;
myTree1.HG2_A10_P.t:=0.000249577;
myTree1.HG2_B1_P.t:=0;
myTree1.HG2_A12_P.t:=0;
myTree1.CHP3_B6_P.t:=0.00636273;
myTree1.HA3_A4.t:=0;
myTree1.HA1_A4_P.t:=0.000248571;
myTree1.Node47.t:=0.0109742;
myTree1.Node55.t:=0.00391241;
myTree1.Node54.t:=0.0010105;
myTree1.BJ4_A3_P.t:=0.00137033;
myTree1.Node7.t:=0.00212069;
myTree1.CB2_A7.t:=0.000751243;
myTree1.Node12.t:=0;
myTree1.CDM4_A3_P.t:=0.00050051;
myTree1.Node11.t:=0.000509885;
myTree1.BS7_A5.t:=0.00323298;
myTree1.Node10.t:=0.00117747;
myTree1.CHP2_A5.t:=0.00150956;
myTree1.Node13.t:=0;
myTree1.Node14.t:=0.000503596;
myTree1.CA3_B3_P.t:=0.000754232;
myTree1.CA1_A8.t:=0.00125559;
myTree1.CB1_A12.t:=0.000501086;
myTree1.Node17.t:=0.000502439;
myTree1.CD7b_A11_P.t:=0.00024925;
myTree1.Node9.t:=0.0039536;
myTree1.Node31.t:=0.000418098;
myTree1.Node33.t:=0.000515897;
myTree1.CDM6_A8.t:=0.000250097;
myTree1.Node29.t:=0.000125062;
myTree1.Node25.t:=0.00792167;
myTree1.Node27.t:=0.00041945;
myTree1.CDM6_A5.t:=0;
myTree1.CHP2_A10_P.t:=0.0026468;
myTree1.Node8.t:=0.00335462;

University of Cape Town

myTree1.HG2_A9_P.t:=0.00421091;
myTree1.GC8_A11.t:=0.0022704;
myTree1.CAN3_A6.t:=0.00150208;
myTree1.HA3_A3.t:=0.00276997;
myTree1.Node66.t:=0;
myTree1.CD7b_B5_P.t:=0.00281446;
myTree1.CA1_A3_P.t:=0.00679279;
myTree1.CHH10_A3_P.t:=0.00496893;
myTree1.CDM8_A9_P.t:=0.00165468;
myTree1.Node110.t:=0.00101109;
myTree1.CD7_A8_P.t:=0.00164329;
myTree1.CB1_A10_P.t:=0.000643581;
myTree1.Node97.t:=0.00174671;
myTree1.GC10_A12_P.t:=0.000251826;
myTree1.CHP3_A12.t:=0.00802465;
myTree1.HA1_A10.t:=0.00862744;
myTree1.GC10_A2_P.t:=0.000253429;
myTree1.CHP3_A10_P.t:=0.00644101;
myTree1.Node98.t:=0.00437395;
myTree1.Node108.t:=0.00167792;
myTree1.Node2.t:=0.00390231;
myTree1.Node72.t:=0.0135382;
myTree1.Node88.t:=0.00288093;
myTree1.CAN3_A9_P.t:=0.000503784;
myTree1.Node119.t:=0.00050282;
myTree1.CM6_A12.t:=0.000503392;
myTree1.Node90.t:=0.000925552;
myTree1.Node102.t:=0.0021355;
myTree1.Node104.t:=0.00102634;
myTree1.Node106.t:=0;
myTree1.Node96.t:=0.00281281;
myTree1.Node92.t:=0.00576115;
myTree1.Node94.t:=0.000663541;
myTree1.HG4_A9_P.t:=0.0163667;
myTree1.Node69.t:=0.0151686;
myTree1.HA1_A8.t:=0.00027018;
myTree1.HA1_A6_P.t:=0.000740252;
myTree1.Node3.t:=0.0116403;
myTree1.HG2_B9_P.t:=0.000355084;
myTree1.HG2_A4_P.t:=0.000907729;
myTree1.Node4.t:=0.000346362;
myTree1.Node60.t:=0.0115746;
myTree1.Node62.t:=0.00103971;
myTree1.Node64.t:=0.00024987;
myTree1.Node52.t:=0.00152817;
myTree1.Node40.t:=0.00218827;
myTree1.Node46.t:=0.00146772;
myTree1.Node78.t:=0.000222681;
myTree1.CA3_A8_P.t:=0.00327665;
myTree1.BS7_A10.t:=0.00160172;
myTree1.Node74.t:=0.0073392;
myTree1.Node85.t:=0.0143198;
myTree1.CD7_A7_P.t:=0.0155596;

University of Cape Town

```
myTree1.Node73.t:=0.00390712;
myTree1.HG4_A10_P.t:=0.00739958;
myTree1.HG2_A2_P.t:=0.00175202;
myTree1.Node77.t:=0.00108641;
myTree1.HG4_A12_P.t:=0.00093638;
myTree1.Node76.t:=0.00113484;
myTree1.Node75.t:=0.00548586;
myTree1.HG2_A6_P.t:=0.00178151;

DependentRateMatrix = {
{*,t*p1,W0*tt*a,0,W0*tt*b,0,W0*tt*c,0}
{t*p1,*,0,W1*tt*a,0,W1*tt*b,0,W1*tt*c}
{W0*tt*a,0,*,t*p1,W0*tt*d,0,W0*tt*e,0}
{0,W1*tt*a,t*p1,*,0,W1*tt*d,0,W1*tt*e}
{W0*tt*b,0,W0*tt*d,0,*,t*p1,W0*tt*f,0}
{0,W1*tt*b,0,W1*tt*d,t*p1,*,0,W1*tt*f}
{W0*tt*c,0,W0*tt*e,0,W0*tt*f,0,*,t*p1}
{0,W1*tt*c,0,W1*tt*e,0,W1*tt*f,t*p1,*}
};
Model Dep = (DependentRateMatrix, obsFreqs);
Tree myTree2 = TREE;
global P_1 = Random(0,1);
P_1 :< 1;
bn = BranchName(myTree1,-1);
for(k=0; k< Columns(bn) -1; k=k+1){
outString = "myTree2."+bn[k]+".t := myTree1."+bn[k]+".t";
ExecuteCommands(outString);
outString = "myTree2."+bn[k]+".tt := myTree1."+bn[k]+".tt";
ExecuteCommands(outString);
}
LikelihoodFunction theLikFun = (myFilter, myTree1,myFilter, myTree2,"
Log((1-P_1)*SITE_LIKELIHOOD[0]+P_1*SITE_LIKELIHOOD[1])");
Optimize(results,theLikFun);
fprintf(stdout,theLikFun);
```

University of Cape Town

IV.3 Parameter file used to run Mayrose and Otto's (2011) method

```
_mainType Optimize_Model
_treeFile ../../tree.nwk
_characterFile ../groups.fas
_seqFile ../../ORnoAGSalignment.fas
_outDir RESULTS
_outFile traitRate.res
_logFile log.txt
_scaledTreeFile scaled.tree
_sequenceType nc
_seqModelType K2
_charModelParam1 0.5
_charModelParam2 5
_gammaParam 0.5
_relRate 1
_logValue 3
_bScaleTree 1
_stochasicMappingIterations 200
_treeLength 1.0
```

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

The tangled history of olfaction in African mole-rats, Bathyergidae: insights from Olfactory Receptor genes

Sofia Stathopoulos

Supplementary Material

Supervisors:
Dr Colleen O'Ryan, Department of Molecular and Cell Biology
Dr Jacqueline M Bishop, Department of Zoology

Thesis presented for the degree of
DOCTOR OF PHYLOSOPHY
In the Department of Molecular and Cell Biology

Faculty of Science
UNIVERSITY OF CAPE TOWN
August 2011

Contents

University of Cape Town

Table 1 dN/dS ratios across clades A-D Ratios are indicated at each codon site, for every clade; dN/dS >1 positive selection, dN/dS = 1 neutral evolution, dN/dS <1 purifying selection.

Clade A

Selecton Bayesian dN/dS Results
Displayed on sequence 1

POS	AMINO	dN/dS	[Confidence Interval](* if lower bound > 1)		POSTERIOR PROBABILITIES							
				w =	0.0074	0.037	0.08	0.14	0.21	0.3	0.42	0.63	1.3
1	A	0.12	[0.0074,0.42]	0.23	0.2	0.17	0.14	0.11	0.075	0.047	0.022	0.0032	
2	D	0.1	[0.0074,0.42]	0.26	0.23	0.18	0.13	0.093	0.058	0.031	0.011	0.00066	
3	I	0.081	[0.0074,0.3]	0.3	0.25	0.18	0.12	0.076	0.041	0.018	0.0046	0.00014	
4	G	1.2	[0.42,1.3]	3.4e-10	9.2e-07	3.6e-05	0.00041	0.0026	0.011	0.041	0.14	0.8	
5	F	0.079	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.073	0.038	0.016	0.0039	0.0001	
6	T	0.095	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.089	0.054	0.027	0.0088	0.00045	
7	S	0.094	[0.0074,0.42]	0.28	0.23	0.18	0.13	0.089	0.053	0.027	0.0087	0.00045	
8	T	0.23	[0.037,0.63]	0.021	0.09	0.15	0.19	0.19	0.17	0.12	0.059	0.0066	
9	T	0.5	[0.08,1.3]	0.00037	0.0082	0.032	0.073	0.13	0.18	0.22	0.22	0.13	
10	V	0.72	[0.14,1.3]	6.6e-06	0.00071	0.0059	0.023	0.059	0.12	0.2	0.28	0.31	
11	P	0.13	[0.0074,0.63]	0.22	0.2	0.17	0.14	0.11	0.079	0.051	0.024	0.004	
12	K	0.32	[0.037,1.3]	0.013	0.056	0.1	0.14	0.17	0.18	0.17	0.12	0.043	
13	M	0.08	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.075	0.04	0.017	0.004	8.4e-05	
14	L	0.16	[0.0074,0.63]	0.19	0.18	0.16	0.14	0.12	0.093	0.068	0.041	0.015	
15	V	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0024	
16	N	0.1	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.095	0.061	0.033	0.012	0.00085	
17	I	0.21	[0.0074,0.63]	0.025	0.1	0.17	0.2	0.19	0.16	0.1	0.046	0.0043	
18	Q	0.4	[0.037,1.3]	0.0094	0.043	0.082	0.12	0.15	0.17	0.18	0.15	0.093	
19	T	0.6	[0.08,1.3]	0.00022	0.005	0.021	0.051	0.096	0.15	0.21	0.25	0.22	
20	Q	0.42	[0.037,1.3]	0.0086	0.04	0.077	0.11	0.15	0.17	0.18	0.16	0.11	
21	S	0.78	[0.21,1.3]	4e-06	0.00045	0.0039	0.016	0.045	0.098	0.18	0.29	0.37	
22	K	0.34	[0.037,1.3]	0.012	0.052	0.098	0.14	0.17	0.18	0.17	0.13	0.055	
23	V	0.75	[0.21,1.3]	2.2e-07	0.00011	0.0019	0.011	0.04	0.1	0.2	0.31	0.33	

24	I	0.098	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.091	0.056	0.029	0.01	0.00064
25	S	1.1	[0.42,1.3]	4.7e-10	1.2e-06	4.8e-05	0.00054	0.0033	0.014	0.049	0.16	0.78
26	Y	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.042	0.018	0.0021
27	A	1.3	[0.63,1.3]	5.4e-13	7.4e-09	6.7e-07	1.4e-05	0.00014	0.001	0.0059	0.038	0.95
28	G	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.042	0.018	0.002
29	C	0.26	[0.037,0.63]	0.017	0.072	0.13	0.17	0.19	0.18	0.15	0.087	0.016
30	I	0.082	[0.0074,0.3]	0.3	0.25	0.18	0.12	0.077	0.042	0.018	0.0049	0.00015
31	T	0.1	[0.0074,0.42]	0.26	0.23	0.18	0.13	0.093	0.058	0.031	0.011	0.00072
32	Q	0.17	[0.0074,0.63]	0.19	0.18	0.16	0.14	0.12	0.094	0.069	0.042	0.014
33	M	0.44	[0.14,1.3]	2.4e-05	0.0024	0.018	0.061	0.13	0.22	0.27	0.24	0.056
34	Y	0.75	[0.14,1.3]	5.5e-06	0.00059	0.005	0.02	0.052	0.11	0.19	0.28	0.34
35	F	0.2	[0.0074,0.63]	0.026	0.11	0.17	0.2	0.19	0.16	0.1	0.042	0.0031
36	F	0.33	[0.08,0.63]	0.00097	0.02	0.068	0.14	0.2	0.23	0.21	0.13	0.021
37	L	0.8	[0.21,1.3]	4.5e-06	0.00049	0.0042	0.017	0.045	0.096	0.17	0.27	0.39
38	L	1	[0.21,1.3]	1.3e-06	0.00015	0.0014	0.006	0.018	0.045	0.098	0.21	0.62
39	F	0.21	[0.0074,0.63]	0.025	0.1	0.17	0.2	0.19	0.16	0.1	0.044	0.0035
40	G	1.1	[0.42,1.3]	1.6e-08	9e-06	0.00017	0.0012	0.0054	0.018	0.052	0.15	0.78
41	E	1.3	[0.63,1.3]	1.2e-13	1.9e-09	1.9e-07	4.6e-06	5.6e-05	0.00047	0.0034	0.027	0.97
42	L	0.28	[0.037,1.3]	0.015	0.067	0.12	0.16	0.18	0.18	0.15	0.098	0.026
43	D	0.28	[0.037,0.63]	0.015	0.067	0.12	0.16	0.18	0.18	0.15	0.096	0.021
44	N	0.43	[0.08,1.3]	0.00052	0.011	0.042	0.092	0.15	0.2	0.23	0.2	0.081
45	F	0.08	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.075	0.04	0.017	0.0043	0.00012
46	L	0.099	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.092	0.057	0.03	0.01	0.0007
47	L	0.17	[0.0074,0.63]	0.19	0.18	0.16	0.14	0.12	0.093	0.069	0.042	0.015
48	A	0.27	[0.037,0.63]	0.016	0.071	0.13	0.17	0.19	0.18	0.15	0.09	0.019
49	V	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0024
50	M	0.08	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.075	0.04	0.017	0.004	8.4e-05
51	A	0.23	[0.037,0.63]	0.022	0.09	0.15	0.19	0.19	0.17	0.12	0.059	0.0065
52	Y	0.12	[0.0074,0.42]	0.23	0.21	0.17	0.14	0.11	0.074	0.045	0.02	0.0027
53	D	0.1	[0.0074,0.42]	0.26	0.23	0.18	0.14	0.094	0.059	0.031	0.011	0.00071
54	R	0.67	[0.037,1.3]	0.0062	0.028	0.053	0.078	0.098	0.12	0.13	0.15	0.34
55	F	0.36	[0.08,1.3]	0.00077	0.016	0.057	0.12	0.18	0.22	0.22	0.15	0.034
56	V	0.31	[0.037,1.3]	0.014	0.061	0.11	0.15	0.18	0.18	0.16	0.11	0.037
57	A	0.22	[0.037,0.63]	0.022	0.092	0.15	0.19	0.19	0.17	0.12	0.056	0.0059
58	I	0.082	[0.0074,0.3]	0.3	0.25	0.18	0.12	0.077	0.042	0.018	0.0048	0.00015
59	C	0.12	[0.0074,0.42]	0.23	0.21	0.17	0.14	0.1	0.073	0.044	0.019	0.0024
60	H	0.98	[0.3,1.3]	5.7e-08	3.1e-05	0.00056	0.0037	0.015	0.045	0.11	0.24	0.58
61	P	0.1	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.095	0.06	0.032	0.012	0.00079
62	L	0.16	[0.0074,0.63]	0.19	0.18	0.16	0.14	0.12	0.091	0.066	0.04	0.014
63	H	0.29	[0.037,1.3]	0.014	0.064	0.12	0.16	0.18	0.18	0.16	0.1	0.027

64	Y	0.11	[0.0074,0.42]	0.25	0.22	0.18	0.14	0.1	0.067	0.039	0.016	0.0016
65	M	1.2	[0.42,1.3]	6.7e-09	3.9e-06	7.7e-05	0.00058	0.0027	0.01	0.032	0.11	0.85
66	L	0.73	[0.21,1.3]	2.3e-07	0.00012	0.002	0.012	0.041	0.1	0.21	0.32	0.31
67	I	0.33	[0.08,0.63]	0.00093	0.019	0.066	0.13	0.19	0.23	0.21	0.13	0.022
68	M	0.08	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.075	0.04	0.017	0.004	8.4e-05
69	N	0.26	[0.037,0.63]	0.018	0.076	0.13	0.17	0.19	0.18	0.14	0.08	0.014
70	H	0.43	[0.08,1.3]	0.0005	0.011	0.041	0.09	0.15	0.2	0.23	0.2	0.074
71	P	1.3	[1.3,1.3]*	4.8e-18	1.7e-12	7.5e-10	4.8e-08	1.3e-06	2.1e-05	0.00028	0.005	0.99
72	L	0.39	[0.08,1.3]	0.00062	0.013	0.048	0.1	0.17	0.21	0.23	0.18	0.05
73	C	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.072	0.043	0.019	0.0022
74	M	1.3	[0.63,1.3]	2.8e-13	4e-09	3.7e-07	8.1e-06	8.9e-05	0.00067	0.0043	0.031	0.96
75	V	1.1	[0.3,1.3]	2.3e-08	1.3e-05	0.00025	0.0018	0.0077	0.025	0.071	0.18	0.71
76	L	0.62	[0.08,1.3]	0.00023	0.0052	0.021	0.051	0.095	0.15	0.2	0.24	0.24
77	V	0.68	[0.14,1.3]	7.9e-06	0.00084	0.0069	0.026	0.066	0.13	0.21	0.28	0.27
78	F	0.92	[0.21,1.3]	2.5e-06	0.00028	0.0024	0.01	0.029	0.065	0.13	0.23	0.53
79	V	1.2	[0.42,1.3]	2.1e-10	5.8e-07	2.3e-05	0.00027	0.0017	0.0078	0.029	0.11	0.85
80	S	0.41	[0.08,1.3]	0.00056	0.012	0.045	0.097	0.16	0.21	0.23	0.19	0.062
81	W	0.69	[0.14,1.3]	0.00017	0.004	0.016	0.041	0.078	0.13	0.19	0.24	0.3
82	I	0.51	[0.14,1.3]	1.8e-05	0.0018	0.014	0.048	0.11	0.19	0.26	0.26	0.12
83	V	0.86	[0.21,1.3]	1.2e-07	6.4e-05	0.0011	0.007	0.026	0.072	0.16	0.29	0.45
84	S	0.28	[0.037,0.63]	0.015	0.066	0.12	0.16	0.18	0.18	0.15	0.099	0.024
85	I	1.2	[0.42,1.3]	7.8e-12	9.9e-08	8e-06	0.00014	0.0012	0.0069	0.031	0.12	0.84
86	L	0.37	[0.037,1.3]	0.011	0.048	0.091	0.13	0.16	0.17	0.17	0.14	0.076
87	H	0.77	[0.21,1.3]	4.8e-06	0.00053	0.0045	0.018	0.048	0.1	0.18	0.28	0.37
88	A	0.23	[0.037,0.63]	0.021	0.089	0.15	0.19	0.19	0.17	0.12	0.06	0.0068
89	L	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0025
90	L	0.32	[0.037,1.3]	0.013	0.058	0.11	0.15	0.17	0.18	0.16	0.12	0.043
91	Q	0.73	[0.14,1.3]	0.00015	0.0035	0.015	0.037	0.071	0.12	0.18	0.24	0.34
92	S	0.92	[0.3,1.3]	7.8e-08	4.2e-05	0.00075	0.0049	0.019	0.056	0.13	0.27	0.51
93	L	0.3	[0.037,1.3]	0.014	0.062	0.11	0.15	0.18	0.18	0.16	0.11	0.032
94	M	0.08	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.075	0.04	0.017	0.004	8.4e-05
95	V	0.28	[0.037,1.3]	0.015	0.067	0.12	0.16	0.18	0.18	0.15	0.098	0.026
96	L	0.37	[0.037,1.3]	0.011	0.048	0.091	0.13	0.16	0.17	0.17	0.14	0.079
97	Q	1.2	[0.42,1.3]	8.4e-09	4.8e-06	9.4e-05	0.00069	0.0032	0.011	0.035	0.11	0.84
98	L	0.17	[0.0074,0.63]	0.19	0.17	0.16	0.14	0.12	0.094	0.07	0.044	0.016
99	S	0.39	[0.08,1.3]	0.00062	0.013	0.048	0.1	0.17	0.21	0.23	0.18	0.051
100	F	0.08	[0.0074,0.3]	0.31	0.25	0.18	0.12	0.075	0.04	0.017	0.0043	0.00012
101	C	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.042	0.018	0.002
102	T	1	[0.3,1.3]	3.6e-08	1.9e-05	0.00036	0.0025	0.01	0.032	0.085	0.2	0.67
103	D	0.41	[0.08,1.3]	0.00054	0.011	0.043	0.095	0.16	0.21	0.23	0.19	0.063

104	L	0.46	[0.08,1.3]	0.00042	0.0091	0.035	0.08	0.14	0.19	0.23	0.21	0.1
105	K	0.35	[0.037,1.3]	0.011	0.05	0.094	0.13	0.16	0.18	0.17	0.14	0.062
106	I	0.083	[0.0074,0.3]	0.3	0.25	0.18	0.13	0.078	0.042	0.019	0.005	0.00017
107	P	0.14	[0.0074,0.63]	0.21	0.19	0.17	0.14	0.11	0.084	0.057	0.03	0.0066
108	H	0.11	[0.0074,0.42]	0.25	0.22	0.18	0.14	0.1	0.067	0.039	0.016	0.0015
109	F	0.083	[0.0074,0.3]	0.3	0.25	0.18	0.13	0.078	0.042	0.019	0.005	0.00016
110	F	0.2	[0.0074,0.63]	0.028	0.11	0.18	0.21	0.19	0.15	0.092	0.036	0.0023
111	C	0.12	[0.0074,0.42]	0.23	0.21	0.17	0.14	0.11	0.073	0.044	0.02	0.0024
112	E	0.14	[0.0074,0.63]	0.21	0.19	0.17	0.14	0.11	0.084	0.057	0.03	0.0064
113	L	0.098	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.092	0.057	0.029	0.01	0.0006
114	N	0.11	[0.0074,0.42]	0.25	0.22	0.18	0.14	0.099	0.065	0.037	0.014	0.0012
115	Q	0.45	[0.037,1.3]	0.0079	0.037	0.071	0.11	0.14	0.16	0.18	0.17	0.14
116	V	0.85	[0.21,1.3]	3.6e-06	0.0004	0.0034	0.014	0.038	0.083	0.16	0.26	0.45
117	A	0.36	[0.08,1.3]	0.00078	0.016	0.058	0.12	0.18	0.22	0.22	0.15	0.034
118	Q	0.92	[0.21,1.3]	2.6e-06	0.00029	0.0025	0.01	0.029	0.066	0.13	0.23	0.53
119	L	1.2	[0.42,1.3]	2.6e-10	7e-07	2.8e-05	0.00032	0.002	0.0092	0.034	0.12	0.83
120	A	0.37	[0.08,1.3]	0.00068	0.014	0.052	0.11	0.17	0.22	0.22	0.17	0.042
121	C	0.29	[0.037,1.3]	0.015	0.065	0.12	0.16	0.18	0.18	0.16	0.1	0.026
122	S	0.1	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.096	0.062	0.034	0.013	0.001
123	E	0.26	[0.037,0.63]	0.017	0.073	0.13	0.17	0.19	0.18	0.14	0.086	0.017
124	N	0.25	[0.037,0.63]	0.018	0.078	0.14	0.18	0.19	0.18	0.14	0.076	0.013
125	F	0.35	[0.08,1.3]	0.00079	0.016	0.058	0.12	0.18	0.22	0.22	0.15	0.03
126	L	0.95	[0.3,1.3]	7.2e-08	3.8e-05	0.00068	0.0044	0.017	0.051	0.12	0.25	0.55
127	N	0.26	[0.037,0.63]	0.017	0.073	0.13	0.17	0.19	0.18	0.14	0.085	0.016
128	D	0.25	[0.037,0.63]	0.018	0.078	0.14	0.18	0.19	0.18	0.14	0.077	0.012
129	F	1.3	[1.3,1.3]*	1.6e-15	1.1e-10	2.3e-08	8.4e-07	1.4e-05	0.00016	0.0015	0.017	0.98
130	V	0.32	[0.037,1.3]	0.012	0.055	0.1	0.14	0.17	0.18	0.17	0.12	0.044
131	M	0.91	[0.3,1.3]	2.1e-08	1.5e-05	0.00037	0.003	0.015	0.05	0.13	0.3	0.5
132	H	0.3	[0.037,1.3]	0.014	0.062	0.11	0.15	0.18	0.18	0.16	0.11	0.031
133	F	0.73	[0.21,1.3]	2.4e-06	0.00031	0.0032	0.015	0.047	0.11	0.21	0.31	0.31
134	A	0.35	[0.037,1.3]	0.011	0.051	0.096	0.14	0.16	0.18	0.17	0.13	0.06
135	P	0.79	[0.21,1.3]	4.2e-06	0.00046	0.004	0.016	0.044	0.096	0.18	0.28	0.39
136	V	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0024
137	L	0.44	[0.037,1.3]	0.0086	0.039	0.076	0.11	0.14	0.16	0.17	0.16	0.13
138	L	0.64	[0.08,1.3]	0.00022	0.0049	0.02	0.048	0.091	0.14	0.2	0.24	0.26
139	G	1.2	[0.63,1.3]	3.4e-12	4.5e-08	3.8e-06	7.1e-05	0.00065	0.004	0.019	0.091	0.89
140	A	1.2	[0.42,1.3]	2.4e-10	6.5e-07	2.6e-05	0.00031	0.002	0.0091	0.034	0.12	0.83
141	G	0.52	[0.08,1.3]	0.00032	0.0071	0.028	0.066	0.12	0.17	0.22	0.23	0.15
142	S	0.42	[0.08,1.3]	0.00053	0.011	0.042	0.093	0.15	0.21	0.23	0.19	0.067
143	L	0.36	[0.037,1.3]	0.005	0.029	0.069	0.12	0.17	0.2	0.2	0.16	0.048

144	A	0.64	[0.14,1.3]	8.7e-06	0.00092	0.0075	0.028	0.072	0.14	0.23	0.29	0.23
145	G	0.28	[0.037,0.63]	0.015	0.066	0.12	0.16	0.18	0.18	0.15	0.098	0.023
146	I	0.2	[0.0074,0.63]	0.027	0.11	0.17	0.2	0.19	0.15	0.098	0.04	0.0028
147	I	0.42	[0.08,1.3]	0.00051	0.011	0.041	0.091	0.15	0.2	0.23	0.2	0.074
148	Y	0.27	[0.037,0.63]	0.016	0.07	0.13	0.17	0.19	0.18	0.15	0.091	0.02
149	S	0.1	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.095	0.06	0.032	0.012	0.00086
150	Y	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.042	0.018	0.0021
151	S	0.59	[0.14,1.3]	1.1e-05	0.0012	0.0095	0.035	0.085	0.16	0.24	0.29	0.18
152	K	0.13	[0.0074,0.63]	0.22	0.2	0.17	0.14	0.11	0.079	0.051	0.025	0.0044
153	I	0.11	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.096	0.062	0.034	0.013	0.0012
154	V	0.23	[0.037,0.63]	0.021	0.089	0.15	0.19	0.19	0.17	0.12	0.061	0.0078
155	S	0.24	[0.037,0.63]	0.02	0.084	0.14	0.18	0.19	0.17	0.13	0.067	0.0088
156	S	0.6	[0.14,1.3]	1.1e-05	0.0011	0.0092	0.034	0.083	0.16	0.24	0.29	0.19
157	V	0.26	[0.037,0.63]	0.017	0.074	0.13	0.17	0.19	0.18	0.14	0.085	0.018
158	L	1.1	[0.3,1.3]	2.9e-08	1.6e-05	0.0003	0.0021	0.0088	0.028	0.076	0.19	0.69
159	E	0.94	[0.21,1.3]	2.1e-06	0.00024	0.0021	0.0089	0.026	0.06	0.12	0.23	0.56
160	I	0.2	[0.0074,0.63]	0.028	0.11	0.18	0.21	0.19	0.15	0.093	0.036	0.0022
161	S	0.14	[0.0074,0.63]	0.21	0.19	0.17	0.14	0.11	0.083	0.055	0.029	0.0064
162	S	0.16	[0.0074,0.63]	0.19	0.18	0.16	0.14	0.12	0.092	0.067	0.04	0.013
163	A	0.42	[0.08,1.3]	0.00052	0.011	0.042	0.093	0.15	0.21	0.23	0.2	0.07
164	Q	0.97	[0.21,1.3]	1.8e-06	0.0002	0.0018	0.0077	0.023	0.053	0.11	0.21	0.59
165	G	0.15	[0.0074,0.63]	0.2	0.18	0.16	0.14	0.11	0.089	0.063	0.035	0.0093
166	K	0.65	[0.14,1.3]	0.0002	0.0046	0.019	0.046	0.087	0.14	0.2	0.24	0.27
167	F	1.2	[0.42,1.3]	3.4e-10	9.2e-07	3.6e-05	0.00041	0.0026	0.011	0.041	0.14	0.81
168	K	0.14	[0.0074,0.63]	0.21	0.19	0.17	0.14	0.11	0.086	0.058	0.031	0.007
169	A	0.13	[0.0074,0.63]	0.22	0.2	0.17	0.14	0.11	0.081	0.054	0.027	0.0052
170	F	0.085	[0.0074,0.42]	0.3	0.24	0.18	0.13	0.08	0.044	0.02	0.0056	0.00019
171	S	0.094	[0.0074,0.42]	0.28	0.23	0.18	0.13	0.089	0.053	0.027	0.0087	0.00045
172	T	0.096	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.09	0.055	0.028	0.0092	0.00049
173	C	0.12	[0.0074,0.42]	0.23	0.21	0.17	0.14	0.11	0.073	0.044	0.02	0.0024
174	S	1.2	[0.42,1.3]	1.1e-10	3.1e-07	1.3e-05	0.00015	0.001	0.005	0.02	0.084	0.89
175	S	0.1	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.095	0.06	0.033	0.012	0.00088
176	H	0.11	[0.0074,0.42]	0.25	0.22	0.18	0.14	0.1	0.067	0.038	0.016	0.0015
177	L	0.09	[0.0074,0.42]	0.28	0.24	0.18	0.13	0.085	0.05	0.024	0.0072	0.0003
178	S	0.15	[0.0074,0.63]	0.2	0.18	0.16	0.14	0.11	0.089	0.062	0.035	0.0097
179	V	0.097	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.09	0.055	0.028	0.0096	0.00062
180	V	0.091	[0.0074,0.42]	0.28	0.24	0.18	0.13	0.085	0.05	0.024	0.0076	0.0004
181	F	0.24	[0.037,0.63]	0.02	0.085	0.15	0.18	0.19	0.17	0.13	0.066	0.009
182	L	0.42	[0.037,1.3]	0.0099	0.044	0.081	0.12	0.14	0.16	0.17	0.16	0.12
183	F	0.2	[0.0074,0.63]	0.027	0.11	0.18	0.2	0.19	0.15	0.096	0.039	0.0028

184	Y	0.12	[0.0074,0.42]	0.23	0.21	0.17	0.14	0.11	0.074	0.045	0.02	0.0027
185	G	0.75	[0.21,1.3]	5.4e-06	0.00059	0.0049	0.019	0.052	0.11	0.19	0.28	0.34
186	T	0.14	[0.0074,0.63]	0.21	0.19	0.17	0.14	0.11	0.083	0.055	0.029	0.0059
187	G	1.2	[0.42,1.3]	3.1e-10	8.4e-07	3.3e-05	0.00038	0.0024	0.011	0.039	0.14	0.81
188	L	0.19	[0.0074,0.63]	0.18	0.17	0.15	0.14	0.12	0.098	0.076	0.05	0.023
189	G	0.31	[0.037,1.3]	0.013	0.059	0.11	0.15	0.18	0.18	0.16	0.11	0.036
190	V	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0024
191	Y	0.11	[0.0074,0.42]	0.25	0.21	0.18	0.14	0.1	0.068	0.039	0.016	0.0016
192	I	0.57	[0.14,1.3]	1.2e-05	0.0013	0.01	0.037	0.09	0.17	0.25	0.28	0.17
193	G	0.29	[0.037,1.3]	0.015	0.064	0.12	0.16	0.18	0.18	0.16	0.1	0.026
194	S	0.11	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.097	0.063	0.035	0.013	0.0011
195	A	0.096	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.09	0.054	0.028	0.0091	0.0005
196	T	1.2	[0.42,1.3]	2.6e-10	7.1e-07	2.8e-05	0.00032	0.002	0.0089	0.033	0.12	0.84
197	V	0.68	[0.14,1.3]	7.5e-06	0.0008	0.0066	0.025	0.065	0.13	0.21	0.29	0.27
198	H	1.1	[0.3,1.3]	3.1e-08	1.7e-05	0.00032	0.0022	0.0092	0.029	0.078	0.19	0.69
199	S	0.66	[0.14,1.3]	8e-06	0.00085	0.007	0.026	0.068	0.13	0.22	0.29	0.25
200	S	0.15	[0.0074,0.63]	0.2	0.18	0.16	0.14	0.11	0.089	0.063	0.036	0.01
201	H	0.27	[0.037,0.63]	0.016	0.071	0.13	0.17	0.19	0.18	0.15	0.088	0.017
202	S	0.41	[0.037,1.3]	0.009	0.041	0.079	0.12	0.15	0.17	0.18	0.16	0.11
203	S	1.2	[0.63,1.3]	2.4e-12	3.2e-08	2.7e-06	5.2e-05	0.00049	0.0031	0.016	0.079	0.9
204	A	0.97	[0.21,1.3]	1.8e-06	0.00021	0.0019	0.0078	0.023	0.054	0.11	0.22	0.59
205	K	1.3	[1.3,1.3]*	4.1e-15	2.8e-10	5.4e-08	1.9e-06	2.9e-05	0.00029	0.0023	0.021	0.98
206	A	0.098	[0.0074,0.42]	0.27	0.23	0.18	0.13	0.091	0.056	0.029	0.01	0.00062
207	S	0.16	[0.0074,0.63]	0.2	0.18	0.16	0.14	0.12	0.091	0.066	0.039	0.012
208	V	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0024
209	M	0.19	[0.0074,0.63]	0.028	0.11	0.18	0.21	0.19	0.15	0.091	0.034	0.0017
210	Y	0.11	[0.0074,0.42]	0.25	0.22	0.18	0.14	0.1	0.067	0.039	0.016	0.0015
211	T	0.1	[0.0074,0.42]	0.26	0.22	0.18	0.14	0.095	0.06	0.032	0.012	0.00083
212	V	0.12	[0.0074,0.42]	0.24	0.21	0.18	0.14	0.1	0.071	0.043	0.019	0.0024
213	V	0.088	[0.0074,0.42]	0.29	0.24	0.18	0.13	0.082	0.047	0.022	0.0065	0.00029

Clade C

Selecton Bayesian dN/dS Results
Displayed on sequence 1

POS	AMINO	dN/dS	[Confidence Interval](* if lower bound > 1)	POSTERIOR PROBABILITIES								
			w =	0.0037	0.023	0.053	0.096	0.15	0.23	0.34	0.55	1.5
1	A	0.29	[0.0037,1.5]	0.13	0.13	0.13	0.12	0.11	0.11	0.094	0.077	0.1
2	D	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.091	0.07	0.07
3	I	0.2	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.082	0.057	0.046
4	G	0.9	[0.053,1.5]	0.0019	0.011	0.025	0.044	0.065	0.09	0.12	0.15	0.49
5	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043
6	T	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.062	0.05
7	T	0.98	[0.053,1.5]	0.0016	0.0093	0.021	0.037	0.056	0.078	0.1	0.14	0.56
8	T	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.087	0.064	0.055
9	T	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.1	0.091	0.071	0.082
10	M	0.98	[0.053,1.5]	0.0012	0.0074	0.018	0.032	0.052	0.077	0.11	0.15	0.55
11	P	0.35	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.083	0.14
12	K	0.31	[0.0037,1.5]	0.13	0.13	0.12	0.12	0.11	0.1	0.094	0.079	0.11
13	M	0.17	[0.0037,1.5]	0.17	0.16	0.15	0.14	0.12	0.1	0.081	0.054	0.032
14	L	1.1	[0.096,1.5]	0.00095	0.0058	0.014	0.025	0.04	0.059	0.086	0.13	0.64
15	V	1.3	[0.23,1.5]	1.1e-05	0.00038	0.002	0.0062	0.014	0.029	0.056	0.11	0.79
16	N	1.3	[0.23,1.5]	7.8e-06	0.00028	0.0015	0.0047	0.011	0.023	0.045	0.091	0.82
17	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.083	0.059	0.048
18	Q	1.4	[0.34,1.5]	3.7e-06	0.00013	0.00073	0.0023	0.0058	0.013	0.026	0.059	0.89
19	L	1.4	[0.34,1.5]	1.9e-06	7.4e-05	0.00043	0.0015	0.0041	0.0099	0.023	0.058	0.9
20	H	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.074	0.087
21	S	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.091	0.07	0.071
22	K	1	[0.053,1.5]	0.0014	0.0081	0.019	0.033	0.05	0.071	0.097	0.13	0.59
23	S	1.3	[0.23,1.5]	6.1e-06	0.00022	0.0012	0.0037	0.009	0.019	0.038	0.08	0.85
24	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.083	0.059	0.048
25	S	1.5	[0.55,1.5]	1.6e-08	3.5e-06	4.5e-05	0.00026	0.001	0.0032	0.0095	0.032	0.95
26	Y	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.075	0.089
27	T	1.3	[0.23,1.5]	7.2e-06	0.00026	0.0014	0.0043	0.01	0.022	0.044	0.089	0.83
28	G	0.85	[0.053,1.5]	0.0021	0.012	0.028	0.048	0.072	0.098	0.13	0.16	0.45
29	C	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.08
30	L	0.29	[0.0037,1.5]	0.13	0.13	0.13	0.12	0.11	0.1	0.092	0.075	0.1

31	T	0.27	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.1	0.092	0.073	0.088
32	Q	0.38	[0.0037,1.5]	0.12	0.11	0.11	0.11	0.11	0.1	0.095	0.085	0.16
33	I	0.64	[0.053,1.5]	0.0034	0.02	0.044	0.073	0.1	0.13	0.16	0.17	0.29
34	W	1.2	[0.096,1.5]	0.00077	0.0047	0.011	0.02	0.032	0.047	0.069	0.11	0.71
35	F	0.74	[0.053,1.5]	0.0028	0.016	0.036	0.061	0.088	0.12	0.14	0.17	0.37
36	A	1.4	[0.55,1.5]	2.4e-08	5.2e-06	6.6e-05	0.00037	0.0014	0.0044	0.013	0.04	0.94
37	L	1.2	[0.15,1.5]	0.00014	0.0011	0.0038	0.0094	0.019	0.036	0.065	0.12	0.75
38	A	1.4	[0.55,1.5]	2.6e-08	5.6e-06	7.1e-05	0.00039	0.0015	0.0047	0.013	0.042	0.94
39	F	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.048
40	L	1.5	[0.55,1.5]	8.4e-09	1.8e-06	2.4e-05	0.00014	0.00055	0.0018	0.0057	0.021	0.97
41	G	1.4	[0.34,1.5]	3.4e-06	0.00012	0.00069	0.0022	0.0055	0.012	0.026	0.058	0.9
42	L	1.4	[0.34,1.5]	3.8e-06	0.00014	0.00076	0.0024	0.0061	0.013	0.028	0.064	0.89
43	E	1.1	[0.096,1.5]	0.0012	0.007	0.016	0.029	0.044	0.064	0.089	0.13	0.62
44	N	0.84	[0.053,1.5]	0.0022	0.013	0.029	0.05	0.074	0.1	0.13	0.16	0.45
45	G	1.5	[0.55,1.5]	1.1e-08	2.5e-06	3.2e-05	0.00019	0.00074	0.0024	0.0074	0.026	0.96
46	I	1.1	[0.096,1.5]	0.00081	0.0049	0.012	0.022	0.035	0.053	0.078	0.12	0.68
47	L	0.38	[0.0037,1.5]	0.12	0.12	0.11	0.11	0.11	0.1	0.095	0.084	0.16
48	V	1.3	[0.23,1.5]	9.5e-06	0.00034	0.0018	0.0056	0.013	0.027	0.052	0.1	0.8
49	A	1.4	[0.23,1.5]	4.6e-06	0.00016	0.0009	0.0028	0.0069	0.015	0.031	0.067	0.88
50	M	0.17	[0.0037,1.5]	0.17	0.16	0.15	0.14	0.12	0.1	0.081	0.054	0.032
51	A	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.054
52	Y	0.28	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.11	0.11	0.093	0.075	0.093
53	D	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.091	0.07	0.07
54	R	0.96	[0.053,1.5]	0.0016	0.0094	0.022	0.038	0.057	0.08	0.11	0.14	0.54
55	F	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.048
56	V	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.092	0.074	0.09
57	A	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.054
58	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.082	0.058	0.047
59	C	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.075	0.088
60	H	0.89	[0.053,1.5]	0.0019	0.011	0.026	0.044	0.066	0.091	0.12	0.15	0.49
61	P	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.091	0.071	0.072
62	L	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.074	0.087
63	R	0.34	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.082	0.13
64	Y	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.084
65	N	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.091	0.07	0.07
66	V	0.83	[0.053,1.5]	0.0023	0.014	0.03	0.052	0.076	0.1	0.13	0.15	0.44
67	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.083	0.058	0.047
68	M	1.2	[0.15,1.5]	1.3e-05	0.00046	0.0024	0.0073	0.017	0.034	0.062	0.11	0.77
69	N	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.09	0.068	0.065
70	P	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.09	0.068	0.064

71	K	0.32	[0.0037,1.5]	0.13	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.12
72	L	0.87	[0.053,1.5]	0.0021	0.013	0.028	0.048	0.071	0.095	0.12	0.15	0.47
73	C	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.074	0.088
74	W	0.35	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.083	0.14
75	L	1.1	[0.096,1.5]	0.001	0.0061	0.014	0.025	0.039	0.057	0.08	0.11	0.66
76	L	0.37	[0.0037,1.5]	0.12	0.12	0.11	0.11	0.11	0.1	0.095	0.084	0.16
77	V	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.062	0.054
78	L	0.35	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.094	0.082	0.14
79	L	0.85	[0.053,1.5]	0.0021	0.013	0.029	0.049	0.072	0.099	0.13	0.16	0.45
80	S	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.057
81	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043
82	L	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.063	0.052
83	I	0.2	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.082	0.057	0.046
84	S	0.25	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.092	0.071	0.073
85	V	0.21	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.085	0.061	0.052
86	L	0.29	[0.0037,1.5]	0.13	0.13	0.13	0.12	0.11	0.1	0.093	0.076	0.1
87	D	0.86	[0.053,1.5]	0.0021	0.013	0.029	0.049	0.072	0.098	0.13	0.15	0.46
88	A	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.054
89	M	1	[0.053,1.5]	0.0014	0.0081	0.019	0.032	0.05	0.071	0.096	0.13	0.59
90	L	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.088	0.065	0.062
91	H	0.93	[0.053,1.5]	0.0017	0.01	0.023	0.04	0.061	0.085	0.11	0.15	0.52
92	T	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.062	0.05
93	L	1	[0.053,1.5]	0.0013	0.0079	0.018	0.032	0.049	0.071	0.098	0.13	0.59
94	M	1.4	[0.34,1.5]	3.5e-06	0.00013	0.0007	0.0022	0.0055	0.012	0.025	0.057	0.9
95	A	0.98	[0.053,1.5]	0.0015	0.0091	0.021	0.036	0.055	0.078	0.1	0.14	0.56
96	L	0.37	[0.0037,1.5]	0.12	0.12	0.11	0.11	0.11	0.1	0.095	0.084	0.16
97	R	1.1	[0.096,1.5]	0.00093	0.0056	0.013	0.023	0.036	0.053	0.075	0.11	0.68
98	L	0.3	[0.0037,1.5]	0.13	0.13	0.13	0.12	0.11	0.1	0.093	0.076	0.11
99	S	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.091	0.072	0.083
100	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043
101	C	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.08
102	K	1	[0.053,1.5]	0.0014	0.0084	0.019	0.034	0.051	0.073	0.099	0.13	0.58
103	N	0.87	[0.053,1.5]	0.0021	0.012	0.028	0.047	0.07	0.095	0.12	0.15	0.47
104	L	1.1	[0.096,1.5]	0.00096	0.0057	0.013	0.024	0.037	0.054	0.077	0.11	0.67
105	E	0.31	[0.0037,1.5]	0.13	0.13	0.12	0.12	0.11	0.1	0.095	0.079	0.11
106	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.051
107	P	0.81	[0.053,1.5]	0.0024	0.014	0.031	0.053	0.078	0.11	0.13	0.16	0.42
108	H	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.082
109	F	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.048
110	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043

111	C	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.075	0.088
112	E	0.33	[0.0037,1.5]	0.12	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.13
113	L	0.37	[0.0037,1.5]	0.12	0.12	0.11	0.11	0.11	0.1	0.095	0.084	0.16
114	A	1.3	[0.23,1.5]	5.9e-06	0.00021	0.0012	0.0036	0.0088	0.019	0.037	0.079	0.85
115	H	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.075	0.089
116	I	0.84	[0.053,1.5]	0.0022	0.013	0.029	0.049	0.073	0.1	0.13	0.16	0.45
117	L	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.062	0.05
118	K	0.31	[0.0037,1.5]	0.13	0.13	0.12	0.12	0.11	0.1	0.094	0.079	0.11
119	L	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.088	0.065	0.057
120	S	0.8	[0.053,1.5]	0.0024	0.014	0.032	0.054	0.079	0.11	0.13	0.16	0.42
121	C	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.075	0.088
122	S	0.88	[0.053,1.5]	0.0021	0.012	0.027	0.047	0.069	0.095	0.12	0.15	0.47
123	D	0.25	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.092	0.072	0.077
124	I	0.2	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.082	0.057	0.046
125	L	0.86	[0.053,1.5]	0.0022	0.013	0.029	0.049	0.072	0.098	0.13	0.15	0.46
126	M	0.72	[0.053,1.5]	0.0031	0.018	0.04	0.066	0.093	0.12	0.15	0.16	0.35
127	N	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.091	0.07	0.07
128	N	0.86	[0.053,1.5]	0.0021	0.012	0.028	0.048	0.071	0.097	0.12	0.15	0.46
129	I	0.84	[0.053,1.5]	0.0024	0.014	0.031	0.052	0.076	0.1	0.12	0.15	0.45
130	L	0.36	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.083	0.15
131	V	0.99	[0.053,1.5]	0.0015	0.0089	0.021	0.036	0.054	0.076	0.1	0.13	0.57
132	Y	0.28	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.11	0.11	0.093	0.075	0.093
133	V	1.2	[0.15,1.5]	1.3e-05	0.00045	0.0024	0.0073	0.017	0.034	0.063	0.12	0.76
134	V	1.1	[0.096,1.5]	0.0009	0.0055	0.013	0.024	0.039	0.059	0.087	0.13	0.64
135	T	0.33	[0.0037,1.5]	0.13	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.13
136	G	0.95	[0.053,1.5]	0.0016	0.0097	0.022	0.039	0.058	0.082	0.11	0.14	0.54
137	L	1	[0.053,1.5]	0.0014	0.0084	0.019	0.034	0.051	0.073	0.099	0.13	0.58
138	L	1	[0.053,1.5]	0.0015	0.0087	0.02	0.035	0.053	0.075	0.1	0.13	0.57
139	G	0.28	[0.0037,1.5]	0.13	0.13	0.13	0.12	0.11	0.11	0.094	0.076	0.094
140	V	0.73	[0.053,1.5]	0.0029	0.017	0.037	0.063	0.091	0.12	0.15	0.17	0.36
141	V	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.062	0.053
142	P	0.88	[0.053,1.5]	0.002	0.012	0.027	0.046	0.069	0.094	0.12	0.15	0.48
143	L	1.2	[0.15,1.5]	1.2e-05	0.00044	0.0023	0.0071	0.017	0.033	0.062	0.11	0.76
144	S	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.089	0.067	0.063
145	G	0.35	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.083	0.14
146	I	0.24	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.09	0.069	0.071
147	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.051
148	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043
149	S	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.09	0.068	0.064
150	Y	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.084

151	T	0.88	[0.053,1.5]	0.002	0.012	0.027	0.046	0.069	0.093	0.12	0.15	0.48
152	Q	0.39	[0.0037,1.5]	0.11	0.11	0.11	0.11	0.1	0.1	0.095	0.086	0.17
153	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.051
154	V	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.048
155	S	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.089	0.066	0.061
156	S	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.089	0.067	0.063
157	V	0.7	[0.053,1.5]	0.0031	0.018	0.04	0.067	0.096	0.12	0.15	0.17	0.33
158	L	0.32	[0.0037,1.5]	0.13	0.13	0.12	0.12	0.11	0.1	0.094	0.078	0.12
159	K	0.32	[0.0037,1.5]	0.13	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.12
160	I	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.083	0.059	0.048
161	P	1.1	[0.053,1.5]	0.0012	0.0072	0.017	0.029	0.045	0.065	0.09	0.13	0.62
162	S	0.37	[0.0037,1.5]	0.12	0.12	0.11	0.11	0.11	0.1	0.095	0.084	0.15
163	A	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.09	0.067	0.061
164	G	0.95	[0.053,1.5]	0.0017	0.0098	0.022	0.039	0.059	0.082	0.11	0.14	0.53
165	G	0.4	[0.0037,1.5]	0.11	0.11	0.11	0.11	0.1	0.1	0.095	0.086	0.17
166	K	0.3	[0.0037,1.5]	0.13	0.13	0.13	0.12	0.11	0.1	0.094	0.078	0.1
167	Y	0.28	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.11	0.11	0.093	0.075	0.093
168	K	0.32	[0.0037,1.5]	0.13	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.12
169	A	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.056
170	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043
171	S	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.089	0.067	0.062
172	I	1.3	[0.23,1.5]	1.1e-05	0.00038	0.002	0.0062	0.015	0.03	0.056	0.11	0.78
173	C	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.075	0.088
174	V	1.4	[0.23,1.5]	4.8e-06	0.00017	0.00095	0.003	0.0073	0.016	0.032	0.069	0.87
175	S	0.28	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.11	0.1	0.093	0.075	0.095
176	H	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.082
177	L	0.97	[0.053,1.5]	0.0016	0.0096	0.022	0.038	0.057	0.08	0.11	0.14	0.55
178	I	0.24	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.1	0.087	0.066	0.075
179	V	1.2	[0.15,1.5]	1.5e-05	0.00052	0.0028	0.0084	0.019	0.039	0.071	0.13	0.73
180	V	0.73	[0.053,1.5]	0.0029	0.017	0.037	0.063	0.091	0.12	0.15	0.17	0.36
181	S	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.057
182	L	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.093	0.074	0.087
183	F	0.19	[0.0037,1.5]	0.16	0.15	0.15	0.13	0.12	0.1	0.083	0.057	0.043
184	Y	0.28	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.11	0.11	0.093	0.075	0.093
185	G	0.3	[0.0037,1.5]	0.13	0.13	0.12	0.12	0.11	0.1	0.094	0.078	0.11
186	T	0.33	[0.0037,1.5]	0.13	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.13
187	G	0.88	[0.053,1.5]	0.002	0.012	0.027	0.046	0.068	0.094	0.12	0.15	0.48
188	F	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.048
189	G	1	[0.053,1.5]	0.0012	0.0074	0.017	0.03	0.046	0.066	0.092	0.13	0.61
190	V	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.092	0.074	0.09

191	Y	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.084
192	L	0.21	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.086	0.062	0.05
193	S	0.25	[0.0037,1.5]	0.14	0.14	0.13	0.13	0.12	0.11	0.092	0.071	0.073
194	S	0.23	[0.0037,1.5]	0.15	0.14	0.14	0.13	0.12	0.11	0.09	0.068	0.064
195	T	0.33	[0.0037,1.5]	0.13	0.12	0.12	0.12	0.11	0.1	0.095	0.081	0.13
196	G	0.96	[0.053,1.5]	0.0016	0.0095	0.022	0.038	0.057	0.08	0.11	0.14	0.54
197	T	0.79	[0.053,1.5]	0.0025	0.015	0.033	0.055	0.081	0.11	0.14	0.16	0.41
198	L	1.3	[0.23,1.5]	5.5e-06	0.0002	0.0011	0.0034	0.0083	0.018	0.036	0.075	0.86
199	S	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.11	0.088	0.065	0.057
200	S	0.8	[0.053,1.5]	0.0024	0.014	0.032	0.054	0.079	0.11	0.14	0.16	0.41
201	R	0.34	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.082	0.13
202	K	0.31	[0.0037,1.5]	0.13	0.13	0.12	0.12	0.11	0.1	0.094	0.079	0.11
203	N	0.84	[0.053,1.5]	0.0022	0.013	0.029	0.049	0.073	0.1	0.13	0.16	0.45
204	A	0.34	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.082	0.13
205	V	0.9	[0.053,1.5]	0.0019	0.011	0.025	0.044	0.065	0.09	0.12	0.15	0.49
206	A	0.82	[0.053,1.5]	0.0023	0.014	0.031	0.052	0.077	0.1	0.13	0.16	0.43
207	S	0.35	[0.0037,1.5]	0.12	0.12	0.12	0.11	0.11	0.1	0.095	0.082	0.14
208	V	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.092	0.074	0.09
209	M	0.17	[0.0037,1.5]	0.17	0.16	0.15	0.14	0.12	0.1	0.081	0.054	0.032
210	Y	0.26	[0.0037,1.5]	0.14	0.14	0.13	0.12	0.12	0.11	0.093	0.073	0.084
211	T	0.22	[0.0037,1.5]	0.15	0.15	0.14	0.13	0.12	0.1	0.088	0.065	0.056
212	V	0.27	[0.0037,1.5]	0.14	0.13	0.13	0.12	0.12	0.11	0.092	0.074	0.09
213	V	0.2	[0.0037,1.5]	0.16	0.15	0.14	0.13	0.12	0.1	0.084	0.06	0.048

Clade D

Selecton Bayesian dN/dS Results
Displayed on sequence 1

POS	AMINO	dN/dS	[Confidence Interval](* if lower bound > 1)		POSTERIOR PROBABILITIES								
				w =	0.03	0.1	0.18	0.27	0.37	0.48	0.62	0.8	1
1	A	0.19	[0.03,0.8]	0.3	0.23	0.17	0.12	0.079	0.05	0.03	0.014	0.014	
2	D	0.14	[0.03,0.48]	0.38	0.25	0.16	0.097	0.055	0.029	0.014	0.0049	0.0035	
3	I	0.11	[0.03,0.37]	0.48	0.27	0.14	0.068	0.031	0.013	0.0044	0.0011	0.00055	
4	G	0.83	[0.37,1] 7.9e-07	0.00024	0.003	0.014	0.04	0.085	0.15	0.21	0.5		
5	F	0.098	[0.03,0.37]	0.51	0.27	0.13	0.06	0.025	0.0095	0.0031	0.00069	0.00031	
6	T	0.12	[0.03,0.48]	0.42	0.26	0.15	0.084	0.044	0.021	0.0085	0.0026	0.0016	
7	S	0.12	[0.03,0.48]	0.44	0.26	0.15	0.08	0.04	0.018	0.0073	0.0021	0.0012	
8	S	0.53	[0.18,1] 0.00092	0.022	0.074	0.13	0.18	0.19	0.16	0.11	0.13		
9	T	0.82	[0.27,1] 7.4e-06	0.00074	0.0056	0.02	0.047	0.089	0.14	0.2	0.49		
10	V	0.89	[0.48,1] 3.4e-09	1.1e-05	0.00037	0.0033	0.015	0.047	0.11	0.21	0.62		
11	P	0.2	[0.03,0.8]	0.29	0.22	0.17	0.12	0.083	0.055	0.033	0.017	0.018	
12	K	0.94	[0.62,1] 3e-10	1.2e-06	5.2e-05	0.00061	0.0037	0.016	0.052	0.16	0.77		
13	L	0.56	[0.18,1] 0.00072	0.018	0.063	0.12	0.17	0.18	0.17	0.12	0.16		
14	V	0.22	[0.03,0.62]	0.13	0.25	0.24	0.17	0.11	0.058	0.027	0.0087	0.0055	
15	V	0.53	[0.1,1]	0.0048	0.04	0.089	0.13	0.16	0.16	0.15	0.11	0.16	
16	D	0.6	[0.18,1] 0.00051	0.013	0.049	0.099	0.15	0.18	0.17	0.14	0.2		
17	I	0.22	[0.03,0.62]	0.13	0.25	0.24	0.17	0.11	0.058	0.026	0.0087	0.0055	
18	L	0.92	[0.48,1] 9.4e-08	3.4e-05	0.00049	0.0028	0.01	0.03	0.073	0.17	0.71		
19	T	0.84	[0.37,1] 8.3e-08	7.8e-05	0.0015	0.0095	0.033	0.078	0.14	0.22	0.52		
20	H	0.51	[0.1,1]	0.0054	0.044	0.097	0.14	0.17	0.17	0.14	0.1	0.14	
21	S	0.72	[0.27,1] 2.5e-05	0.0022	0.015	0.045	0.09	0.14	0.18	0.19	0.34		
22	R	0.74	[0.27,1] 0.00014	0.0042	0.018	0.045	0.081	0.12	0.16	0.18	0.39		
23	V	0.74	[0.27,1] 2.8e-06	0.00076	0.008	0.032	0.077	0.14	0.19	0.2	0.36		
24	I	0.11	[0.03,0.37]	0.48	0.27	0.14	0.068	0.031	0.013	0.0045	0.0012	0.00058	
25	S	0.56	[0.18,1] 0.00075	0.018	0.064	0.12	0.17	0.18	0.17	0.12	0.16		
26	Y	0.29	[0.03,1] 0.08	0.18	0.2	0.18	0.14	0.099	0.059	0.028	0.025		
27	A	0.98	[0.8,1]	5.2e-15	8.2e-10	2.2e-07	9e-06	0.00015	0.0015	0.012	0.085	0.9	
28	A	0.88	[0.37,1] 1.9e-07	7.1e-05	0.001	0.0058	0.02	0.051	0.11	0.2	0.62		
29	C	0.14	[0.03,0.48]	0.38	0.25	0.16	0.097	0.055	0.029	0.014	0.0049	0.0035	
30	L	0.21	[0.03,0.8]	0.28	0.22	0.16	0.12	0.084	0.056	0.035	0.018	0.02	

31	T	0.53	[0.1,1]		0.0048	0.04	0.089	0.13	0.16	0.16	0.15	0.11	0.16
32	Q	0.26	[0.03,1]	0.23	0.19	0.16	0.12	0.096	0.072	0.05	0.031	0.04	
33	L	0.6	[0.18,1]	8.6e-05	0.0066	0.037	0.094	0.16	0.2	0.19	0.14	0.18	
34	S	0.26	[0.03,0.8]		0.1	0.21	0.22	0.18	0.13	0.081	0.044	0.018	0.014
35	A	0.95	[0.62,1]	2.7e-12	1e-07	1.2e-05	0.00024	0.0022	0.012	0.047	0.16	0.78	
36	F	0.57	[0.18,1]	0.00013	0.0092	0.048	0.11	0.17	0.2	0.18	0.13	0.15	
37	L	0.99	[0.8,1]		6.6e-18	1.2e-11	1.1e-08	9.5e-07	2.9e-05	0.00048	0.0057	0.064	0.93
38	F	0.66	[0.18,1]	0.00037	0.0099	0.037	0.077	0.12	0.15	0.17	0.16	0.27	
39	F	0.095	[0.03,0.37]		0.52	0.27	0.13	0.055	0.022	0.0082	0.0025	0.00053	0.00022
40	G	0.86	[0.37,1]	3.2e-06	0.00035	0.0029	0.011	0.029	0.062	0.11	0.19	0.59	
41	C	0.28	[0.03,0.8]		0.087	0.19	0.21	0.18	0.14	0.092	0.053	0.023	0.019
42	M	0.62	[0.18,1]	5.9e-05	0.0049	0.029	0.08	0.14	0.19	0.2	0.16	0.2	
43	D	0.14	[0.03,0.48]		0.39	0.26	0.16	0.095	0.054	0.028	0.013	0.0046	0.0033
44	S	0.98	[0.8,1]		2.5e-14	3.5e-09	8.1e-07	2.8e-05	0.0004	0.0033	0.02	0.11	0.86
45	M	0.98	[0.8,1]		9e-17	1.3e-10	8.7e-08	5.9e-06	0.00013	0.0016	0.013	0.096	0.89
46	L	0.37	[0.1,1]		0.014	0.098	0.18	0.21	0.19	0.14	0.092	0.043	0.038
47	L	0.27	[0.03,1]	0.23	0.19	0.16	0.12	0.098	0.074	0.053	0.034	0.046	
48	T	0.44	[0.1,1]		0.0088	0.066	0.13	0.17	0.18	0.16	0.12	0.072	0.08
49	V	0.36	[0.03,1]	0.054	0.13	0.17	0.17	0.15	0.12	0.087	0.052	0.061	
50	M	0.2	[0.03,0.62]		0.14	0.27	0.24	0.17	0.1	0.05	0.021	0.0061	0.0032
51	A	0.27	[0.03,0.8]		0.092	0.2	0.22	0.18	0.14	0.088	0.049	0.021	0.017
52	Y	0.14	[0.03,0.48]		0.38	0.25	0.16	0.096	0.055	0.029	0.013	0.0048	0.0035
53	D	0.14	[0.03,0.48]		0.38	0.25	0.16	0.097	0.055	0.029	0.014	0.0049	0.0035
54	R	0.2	[0.03,0.8]		0.29	0.22	0.16	0.12	0.083	0.055	0.034	0.017	0.018
55	F	0.19	[0.03,0.62]		0.16	0.28	0.24	0.16	0.092	0.045	0.018	0.005	0.0026
56	V	0.38	[0.03,1]	0.049	0.12	0.16	0.17	0.15	0.13	0.093	0.058	0.072	
57	A	0.13	[0.03,0.48]		0.42	0.26	0.15	0.086	0.045	0.022	0.0091	0.0029	0.0018
58	I	0.11	[0.03,0.37]		0.48	0.27	0.14	0.068	0.031	0.013	0.0046	0.0012	0.0006
59	C	0.14	[0.03,0.48]		0.38	0.25	0.16	0.097	0.055	0.029	0.014	0.0049	0.0035
60	H	0.49	[0.1,1]		0.0061	0.049	0.11	0.15	0.17	0.17	0.14	0.094	0.12
61	P	0.14	[0.03,0.48]		0.39	0.26	0.16	0.093	0.052	0.026	0.012	0.0041	0.0028
62	L	0.27	[0.03,1]	0.23	0.19	0.16	0.12	0.098	0.074	0.053	0.034	0.046	
63	H	0.63	[0.18,1]	0.0004	0.011	0.041	0.086	0.13	0.17	0.18	0.15	0.24	
64	Y	0.3	[0.03,1]	0.078	0.18	0.2	0.18	0.14	0.1	0.061	0.029	0.026	
65	V	0.96	[0.62,1]	1.2e-10	5.1e-07	2.4e-05	0.00031	0.0021	0.01	0.038	0.14	0.81	
66	V	0.89	[0.48,1]	3e-09	9.6e-06	0.00034	0.0031	0.014	0.044	0.11	0.21	0.63	
67	I	0.22	[0.03,0.62]		0.13	0.25	0.24	0.17	0.11	0.059	0.027	0.0088	0.0055
68	M	0.1	[0.03,0.37]		0.49	0.27	0.13	0.063	0.027	0.01	0.0034	0.00074	0.00032
69	N	0.71	[0.27,1]	2.8e-05	0.0025	0.016	0.048	0.096	0.15	0.18	0.19	0.32	
70	P	0.27	[0.03,0.8]		0.092	0.2	0.22	0.18	0.14	0.088	0.049	0.021	0.017

71	H	0.97	[0.62,1]	2.4e-12	4.3e-08	4.2e-06	8.9e-05	0.00088	0.0056	0.028	0.12	0.84	
72	R	0.79	[0.27,1]	1.5e-06	0.00042	0.0048	0.021	0.055	0.11	0.16	0.21	0.44	
73	C	0.29	[0.03,0.8]		0.081	0.18	0.21	0.18	0.14	0.097	0.058	0.027	0.024
74	Y	0.94	[0.62,1]	3.6e-11	4.3e-07	3e-05	0.00045	0.0033	0.015	0.054	0.16	0.76	
75	L	0.73	[0.27,1]	2.4e-05	0.0021	0.014	0.043	0.087	0.14	0.18	0.19	0.35	
76	L	0.34	[0.03,1]	0.063	0.15	0.18	0.18	0.15	0.11	0.077	0.043	0.047	
77	L	0.37	[0.1,1]		0.014	0.097	0.17	0.21	0.19	0.15	0.092	0.044	0.037
78	L	0.63	[0.18,1]	0.00042	0.011	0.042	0.087	0.13	0.17	0.17	0.15	0.24	
79	L	0.86	[0.37,1]	3.6e-06	0.00038	0.0031	0.012	0.03	0.063	0.11	0.19	0.59	
80	S	0.51	[0.1,1]		0.0057	0.046	0.099	0.14	0.16	0.16	0.14	0.1	0.14
81	V	0.55	[0.18,1]	0.00089	0.021	0.07	0.13	0.17	0.18	0.16	0.12	0.15	
82	F	0.46	[0.1,1]		0.0016	0.035	0.11	0.18	0.21	0.19	0.14	0.075	0.07
83	V	0.83	[0.37,1]	1e-07	9.6e-05	0.0018	0.011	0.036	0.084	0.15	0.22	0.5	
84	S	0.73	[0.27,1]	2.3e-05	0.002	0.014	0.042	0.085	0.14	0.18	0.19	0.35	
85	V	0.6	[0.18,1]	8.6e-05	0.0066	0.037	0.093	0.15	0.19	0.19	0.14	0.18	
86	L	0.82	[0.37,1]	8.5e-07	0.00026	0.0031	0.014	0.041	0.086	0.15	0.21	0.5	
87	D	0.75	[0.27,1]	2e-05	0.0018	0.012	0.037	0.077	0.13	0.17	0.19	0.38	
88	S	0.26	[0.03,0.8]		0.1	0.21	0.22	0.18	0.13	0.081	0.044	0.018	0.014
89	Q	0.5	[0.03,1]	0.027	0.076	0.11	0.13	0.14	0.14	0.12	0.099	0.16	
90	L	0.96	[0.62,1]	1.3e-10	5.6e-07	2.7e-05	0.00034	0.0023	0.011	0.04	0.14	0.81	
91	Q	0.52	[0.1,1]		0.0052	0.042	0.094	0.14	0.16	0.16	0.14	0.11	0.14
92	N	0.59	[0.18,1]	0.00053	0.014	0.051	0.1	0.15	0.18	0.17	0.14	0.2	
93	F	0.83	[0.37,1]	7.6e-07	0.00024	0.0029	0.013	0.039	0.083	0.14	0.21	0.51	
94	T	0.75	[0.27,1]	2.7e-06	0.00072	0.0076	0.03	0.073	0.13	0.18	0.2	0.38	
95	A	0.56	[0.18,1]	0.0007	0.018	0.062	0.12	0.16	0.18	0.17	0.12	0.16	
96	L	0.32	[0.03,1]	0.067	0.16	0.19	0.18	0.15	0.11	0.072	0.038	0.038	
97	Q	0.68	[0.18,1]	0.0017	0.016	0.041	0.072	0.1	0.13	0.15	0.16	0.32	
98	V	0.51	[0.1,1]		0.001	0.024	0.082	0.15	0.19	0.19	0.16	0.098	0.11
99	T	0.26	[0.03,0.8]		0.096	0.2	0.22	0.18	0.13	0.085	0.046	0.019	0.015
100	C	0.29	[0.03,0.8]		0.082	0.18	0.21	0.18	0.14	0.097	0.057	0.027	0.024
101	F	0.21	[0.03,0.62]		0.14	0.26	0.24	0.17	0.1	0.054	0.023	0.0072	0.0042
102	K	0.93	[0.48,1]	5.7e-09	6.8e-06	0.00018	0.0014	0.0068	0.023	0.066	0.17	0.73	
103	D	0.72	[0.27,1]	2.6e-05	0.0023	0.015	0.045	0.092	0.14	0.18	0.19	0.33	
104	V	0.36	[0.03,1]	0.056	0.14	0.17	0.17	0.15	0.12	0.085	0.05	0.057	
105	E	0.89	[0.37,1]	2.2e-07	7.7e-05	0.0011	0.0057	0.019	0.049	0.1	0.19	0.63	
106	I	0.21	[0.03,0.62]		0.14	0.26	0.24	0.17	0.1	0.052	0.023	0.007	0.0041
107	A	0.26	[0.03,0.8]		0.099	0.21	0.22	0.18	0.13	0.082	0.044	0.018	0.013
108	T	0.61	[0.18,1]	0.00047	0.012	0.046	0.094	0.14	0.17	0.17	0.14	0.22	
109	F	0.098	[0.03,0.37]		0.51	0.27	0.13	0.059	0.025	0.0095	0.0031	0.00069	0.00031
110	F	0.096	[0.03,0.37]		0.51	0.27	0.13	0.057	0.023	0.0087	0.0027	0.0006	0.00026

111	C	0.14	[0.03,0.48]		0.39	0.26	0.16	0.094	0.052	0.027	0.012	0.0042	0.0029
112	E	0.7	[0.18,1]	0.00019	0.0057	0.024	0.058	0.1	0.14	0.17	0.18	0.32	
113	T	0.54	[0.18,1]	0.00083	0.02	0.069	0.13	0.17	0.19	0.17	0.11	0.14	
114	S	0.66	[0.18,1]	4.8e-05	0.004	0.024	0.066	0.12	0.17	0.19	0.17	0.26	
115	K	0.79	[0.27,1]	8.3e-05	0.0026	0.012	0.032	0.061	0.1	0.14	0.19	0.46	
116	L	0.61	[0.18,1]	0.00062	0.015	0.05	0.097	0.14	0.17	0.17	0.14	0.21	
117	L	0.59	[0.18,1]	9.4e-05	0.0071	0.039	0.098	0.16	0.2	0.19	0.14	0.17	
118	D	0.74	[0.27,1]	2.1e-05	0.0019	0.013	0.039	0.082	0.13	0.18	0.19	0.37	
119	L	0.49	[0.1,1]		0.0013	0.029	0.091	0.16	0.19	0.19	0.15	0.091	0.097
120	S	0.58	[0.1,1]		0.0036	0.031	0.073	0.11	0.14	0.16	0.15	0.13	0.2
121	C	0.14	[0.03,0.48]		0.39	0.26	0.16	0.093	0.051	0.026	0.012	0.004	0.0027
122	S	0.55	[0.18,1]	0.00073	0.018	0.064	0.12	0.17	0.19	0.17	0.12	0.15	
123	D	0.72	[0.27,1]	2.4e-05	0.0022	0.014	0.044	0.089	0.14	0.18	0.19	0.34	
124	T	0.62	[0.18,1]	0.0004	0.011	0.042	0.09	0.14	0.17	0.18	0.15	0.22	
125	F	0.91	[0.48,1]	1e-08	1.2e-05	0.00029	0.0022	0.01	0.032	0.082	0.19	0.69	
126	F	0.65	[0.18,1]	5.7e-05	0.0046	0.027	0.072	0.13	0.17	0.19	0.16	0.25	
127	K	0.62	[0.1,1]		0.0025	0.023	0.057	0.095	0.13	0.15	0.16	0.14	0.24
128	S	0.73	[0.27,1]	2.4e-05	0.0021	0.014	0.043	0.088	0.14	0.18	0.19	0.35	
129	I	0.9	[0.48,1]	1.9e-08	2e-05	0.00046	0.0033	0.014	0.04	0.094	0.19	0.65	
130	V	0.55	[0.18,1]	0.00072	0.018	0.063	0.12	0.17	0.19	0.17	0.12	0.15	
131	T	0.88	[0.37,1]	2.4e-07	8.3e-05	0.0011	0.006	0.02	0.05	0.1	0.2	0.62	
132	Y	0.29	[0.03,1]	0.079	0.18	0.2	0.18	0.14	0.1	0.06	0.029	0.026	
133	M	0.93	[0.48,1]	9e-11	9.8e-07	6.2e-05	0.00086	0.0056	0.023	0.071	0.18	0.72	
134	F	0.88	[0.48,1]	3.9e-09	1.2e-05	0.00042	0.0037	0.017	0.05	0.11	0.21	0.6	
135	G	0.68	[0.18,1]	0.00026	0.0073	0.03	0.066	0.11	0.15	0.17	0.17	0.3	
136	I	0.91	[0.48,1]	1.4e-09	4.9e-06	0.00019	0.0019	0.0097	0.034	0.089	0.2	0.67	
137	L	0.66	[0.18,1]	0.00035	0.0095	0.037	0.077	0.12	0.16	0.17	0.16	0.27	
138	F	0.33	[0.1,0.8]		0.02	0.12	0.2	0.22	0.18	0.13	0.072	0.03	0.024
139	G	0.16	[0.03,0.62]		0.35	0.25	0.16	0.1	0.063	0.036	0.018	0.0073	0.0059
140	F	0.45	[0.1,1]		0.0018	0.038	0.11	0.18	0.21	0.19	0.13	0.07	0.065
141	L	0.68	[0.27,1]	6.3e-06	0.0016	0.015	0.053	0.11	0.17	0.2	0.18	0.26	
142	P	0.14	[0.03,0.48]		0.38	0.25	0.16	0.096	0.054	0.029	0.013	0.0047	0.0033
143	M	0.89	[0.48,1]	1.7e-09	6.1e-06	0.00024	0.0025	0.013	0.042	0.11	0.21	0.63	
144	S	0.6	[0.1,1]		0.003	0.027	0.064	0.1	0.13	0.15	0.15	0.13	0.23
145	G	0.78	[0.27,1]	8.9e-05	0.0028	0.013	0.033	0.064	0.1	0.15	0.19	0.45	
146	I	0.35	[0.1,1]		0.017	0.11	0.19	0.21	0.19	0.14	0.083	0.038	0.032
147	I	0.47	[0.1,1]		0.0014	0.032	0.099	0.17	0.2	0.19	0.15	0.083	0.085
148	F	0.22	[0.03,0.62]		0.12	0.24	0.23	0.18	0.11	0.064	0.03	0.01	0.0066
149	S	0.12	[0.03,0.48]		0.44	0.26	0.15	0.078	0.038	0.017	0.0067	0.0019	0.0011
150	Y	0.15	[0.03,0.62]		0.37	0.25	0.16	0.1	0.058	0.032	0.015	0.0058	0.0044

151	Y	0.66	[0.18,1]	0.00033	0.0089	0.035	0.075	0.12	0.16	0.17	0.16	0.27	
152	K	0.41	[0.03,1]	0.042	0.11	0.15	0.16	0.15	0.13	0.1	0.068	0.089	
153	I	0.1	[0.03,0.37]		0.49	0.27	0.13	0.063	0.027	0.011	0.0036	0.00085	0.0004
154	V	0.6	[0.18,1]	8.5e-05	0.0065	0.037	0.092	0.15	0.19	0.19	0.14	0.19	
155	S	0.37	[0.1,1]		0.015	0.099	0.18	0.21	0.19	0.14	0.09	0.043	0.037
156	A	0.89	[0.48,1]	2.9e-09	9.3e-06	0.00033	0.003	0.014	0.044	0.11	0.21	0.63	
157	L	0.91	[0.48,1]	2e-10	2e-06	0.00012	0.0015	0.0086	0.032	0.088	0.2	0.67	
158	L	0.47	[0.03,1]	0.032	0.088	0.12	0.14	0.14	0.13	0.12	0.088	0.13	
159	N	0.72	[0.18,1]	0.00018	0.0052	0.022	0.052	0.091	0.13	0.17	0.18	0.35	
160	S	0.49	[0.1,1]		0.0012	0.028	0.09	0.15	0.19	0.19	0.15	0.093	0.1
161	P	0.59	[0.1,1]		0.0031	0.027	0.065	0.1	0.14	0.15	0.15	0.13	0.22
162	S	0.4	[0.03,1]	0.045	0.12	0.15	0.16	0.15	0.13	0.098	0.064	0.083	
163	S	0.77	[0.27,1]	1.3e-05	0.0012	0.0088	0.029	0.065	0.11	0.16	0.2	0.42	
164	V	0.87	[0.37,1]	3.6e-07	0.00012	0.0016	0.008	0.025	0.06	0.12	0.2	0.59	
165	G	0.24	[0.03,1]	0.25	0.2	0.16	0.12	0.092	0.066	0.045	0.026	0.031	
166	R	0.41	[0.03,1]	0.043	0.11	0.15	0.16	0.15	0.13	0.1	0.067	0.087	
167	Y	0.74	[0.27,1]	2.1e-05	0.0019	0.013	0.04	0.082	0.13	0.17	0.19	0.37	
168	K	0.2	[0.03,0.8]		0.3	0.23	0.17	0.12	0.08	0.052	0.031	0.015	0.015
169	T	0.26	[0.03,0.8]		0.097	0.21	0.22	0.18	0.13	0.084	0.046	0.019	0.015
170	F	0.2	[0.03,0.62]		0.16	0.28	0.24	0.16	0.093	0.046	0.019	0.0054	0.003
171	S	0.12	[0.03,0.48]		0.45	0.26	0.15	0.077	0.037	0.017	0.0064	0.0018	0.00097
172	T	0.13	[0.03,0.48]		0.42	0.26	0.15	0.085	0.044	0.021	0.0087	0.0027	0.0016
173	C	0.14	[0.03,0.48]		0.39	0.26	0.16	0.094	0.052	0.027	0.012	0.0042	0.0029
174	S	0.89	[0.48,1]	2.4e-08	2.6e-05	0.00058	0.0041	0.016	0.046	0.1	0.2	0.63	
175	S	0.24	[0.03,0.62]		0.12	0.23	0.23	0.18	0.12	0.07	0.034	0.013	0.0087
176	H	0.16	[0.03,0.62]		0.36	0.25	0.16	0.1	0.063	0.035	0.018	0.0071	0.0057
177	L	0.5	[0.03,1]	0.026	0.074	0.11	0.13	0.14	0.13	0.12	0.1	0.17	
178	S	0.72	[0.18,1]	0.00017	0.005	0.021	0.051	0.089	0.13	0.16	0.18	0.36	
179	V	0.11	[0.03,0.37]		0.47	0.27	0.14	0.07	0.033	0.014	0.005	0.0013	0.00068
180	V	0.11	[0.03,0.37]		0.47	0.27	0.14	0.07	0.033	0.014	0.005	0.0013	0.00068
181	C	0.14	[0.03,0.48]		0.38	0.25	0.16	0.097	0.055	0.029	0.014	0.0049	0.0035
182	L	0.15	[0.03,0.62]		0.37	0.25	0.16	0.1	0.059	0.033	0.016	0.0064	0.005
183	F	0.095	[0.03,0.37]		0.52	0.27	0.13	0.055	0.022	0.0082	0.0025	0.00053	0.00022
184	Y	0.45	[0.1,1]		0.0082	0.063	0.13	0.17	0.18	0.16	0.13	0.076	0.086
185	G	0.25	[0.03,1]	0.24	0.2	0.16	0.12	0.095	0.07	0.049	0.03	0.038	
186	T	0.76	[0.27,1]	0.00012	0.0036	0.016	0.04	0.074	0.11	0.15	0.19	0.41	
187	G	0.76	[0.27,1]	1.5e-05	0.0014	0.01	0.032	0.07	0.12	0.17	0.2	0.41	
188	I	0.92	[0.48,1]	1.4e-10	1.5e-06	9.1e-05	0.0012	0.0073	0.029	0.083	0.2	0.68	
189	G	0.5	[0.03,1]	0.028	0.078	0.11	0.13	0.14	0.13	0.12	0.098	0.16	
190	T	0.74	[0.27,1]	0.00016	0.0046	0.02	0.047	0.083	0.12	0.16	0.18	0.39	

191	Y	0.14	[0.03,0.48]		0.38	0.25	0.16	0.098	0.056	0.03	0.014	0.0052	0.0038
192	L	0.66	[0.18,1]	4.4e-05	0.0037	0.023	0.065	0.12	0.17	0.19	0.17	0.25	
193	G	0.64	[0.18,1]	0.00036	0.0098	0.038	0.081	0.13	0.16	0.17	0.16	0.25	
194	S	0.21	[0.03,0.8]		0.28	0.22	0.16	0.12	0.084	0.057	0.036	0.019	0.021
195	S	0.98	[0.8,1]		8e-17	1.1e-10	7.7e-08	5.3e-06	0.00012	0.0015	0.013	0.093	0.89
196	A	0.81	[0.27,1]	9.3e-06	0.0009	0.0066	0.023	0.052	0.095	0.15	0.2	0.48	
197	S	0.43	[0.03,1]	0.039	0.1	0.14	0.15	0.15	0.13	0.11	0.074	0.1	
198	Y	0.89	[0.48,1]	2.5e-08	2.7e-05	0.0006	0.0042	0.017	0.047	0.11	0.2	0.62	
199	S	0.24	[0.03,0.62]		0.11	0.23	0.23	0.18	0.12	0.07	0.035	0.013	0.0088
200	P	0.59	[0.18,1]	0.00056	0.014	0.053	0.1	0.15	0.18	0.17	0.14	0.19	
201	R	0.83	[0.37,1]	6e-06	0.00061	0.0048	0.017	0.042	0.082	0.14	0.2	0.52	
202	K	0.73	[0.18,1]	0.00017	0.005	0.021	0.05	0.088	0.13	0.16	0.18	0.36	
203	G	0.5	[0.1,1]		0.0057	0.046	0.1	0.15	0.17	0.17	0.14	0.099	0.13
204	M	0.45	[0.1,1]		0.0017	0.037	0.11	0.18	0.21	0.19	0.14	0.07	0.063
205	V	0.35	[0.03,1]	0.058	0.14	0.17	0.17	0.15	0.12	0.083	0.048	0.054	
206	A	0.55	[0.18,1]	0.0008	0.02	0.067	0.13	0.17	0.19	0.17	0.12	0.14	
207	S	0.2	[0.03,0.8]		0.29	0.22	0.16	0.12	0.082	0.054	0.033	0.017	0.018
208	L	0.49	[0.03,1]	0.028	0.079	0.11	0.13	0.14	0.13	0.12	0.096	0.15	
209	M	0.1	[0.03,0.37]		0.49	0.27	0.13	0.063	0.027	0.01	0.0034	0.00074	0.00032
210	Y	0.15	[0.03,0.62]		0.37	0.25	0.16	0.1	0.059	0.032	0.016	0.0059	0.0045
211	T	0.12	[0.03,0.48]		0.44	0.26	0.15	0.08	0.04	0.018	0.0071	0.0021	0.0012
212	V	0.17	[0.03,0.62]		0.33	0.24	0.17	0.11	0.071	0.043	0.024	0.011	0.0098
213	V	0.11	[0.03,0.37]		0.45	0.26	0.15	0.075	0.036	0.016	0.0062	0.0017	0.00096

Figure 1: Nucleotide alignment of all Bathyergidae OR sequences Abbreviations correspond to gene names as per Appendix I.2.

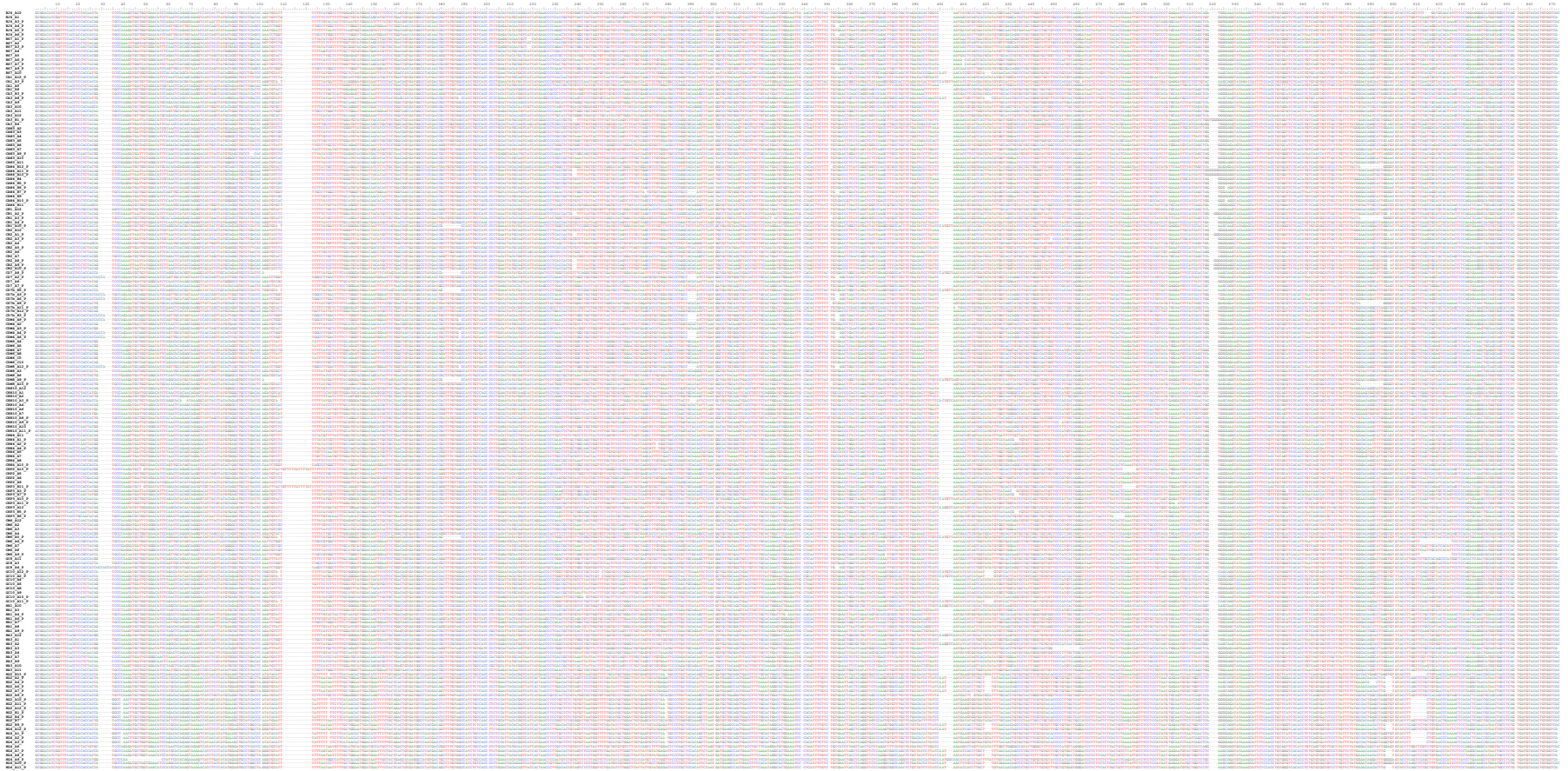


Figure 1 (cont.)

Back to overview

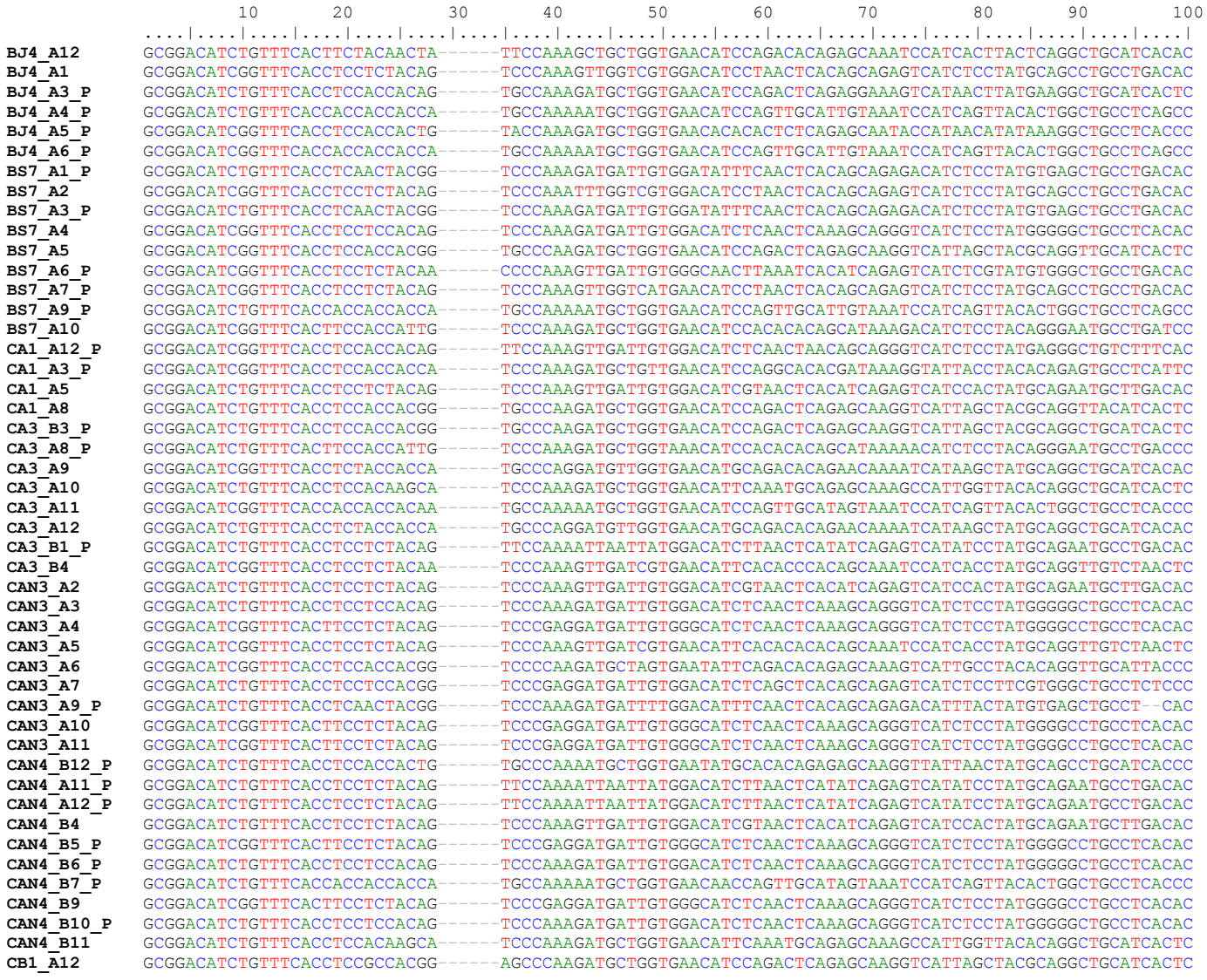


Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

↩ Back to overview

Figure 1 (cont.)

↩ Back to overview

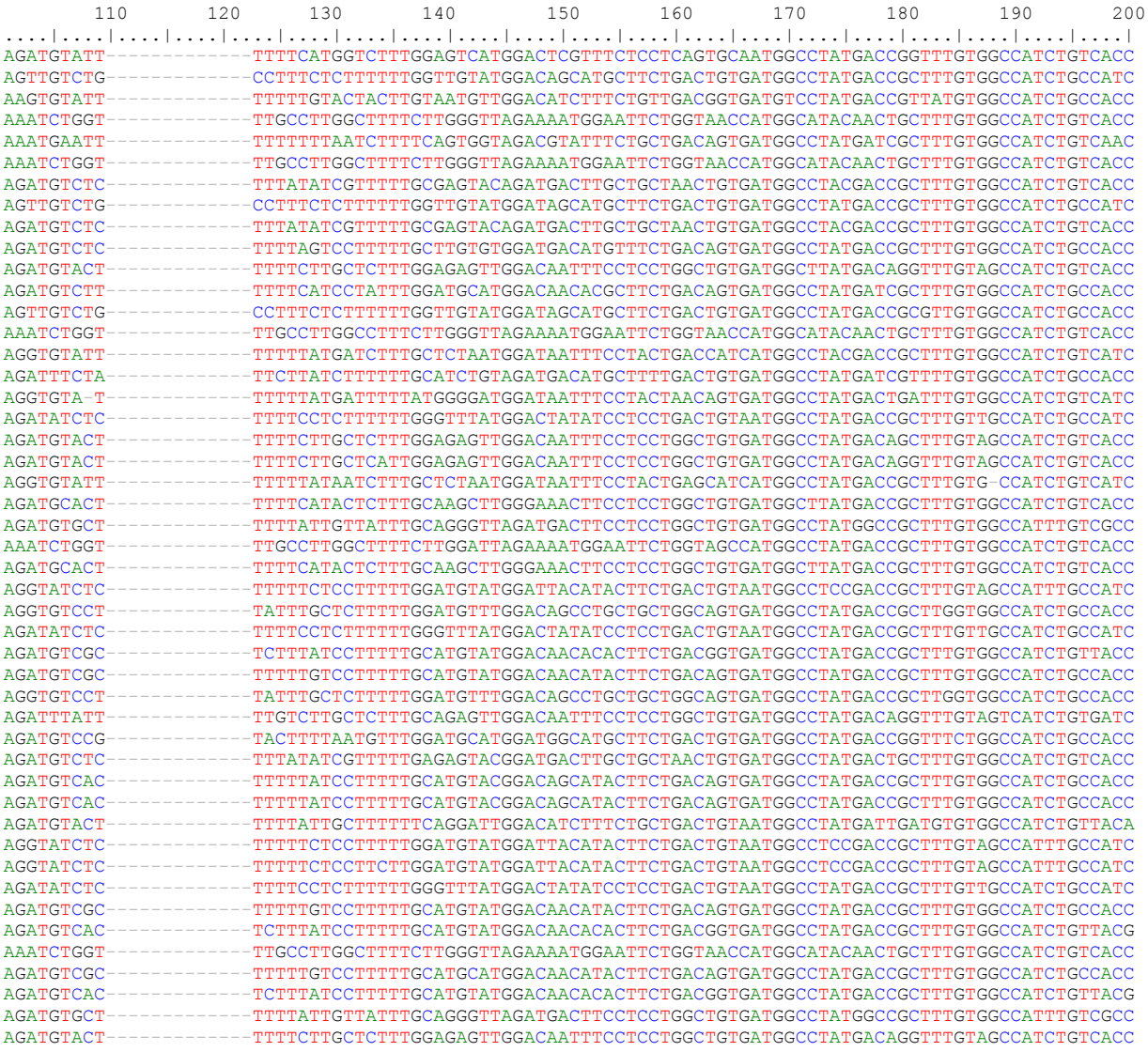


Figure 1 (cont.)

Back to overview

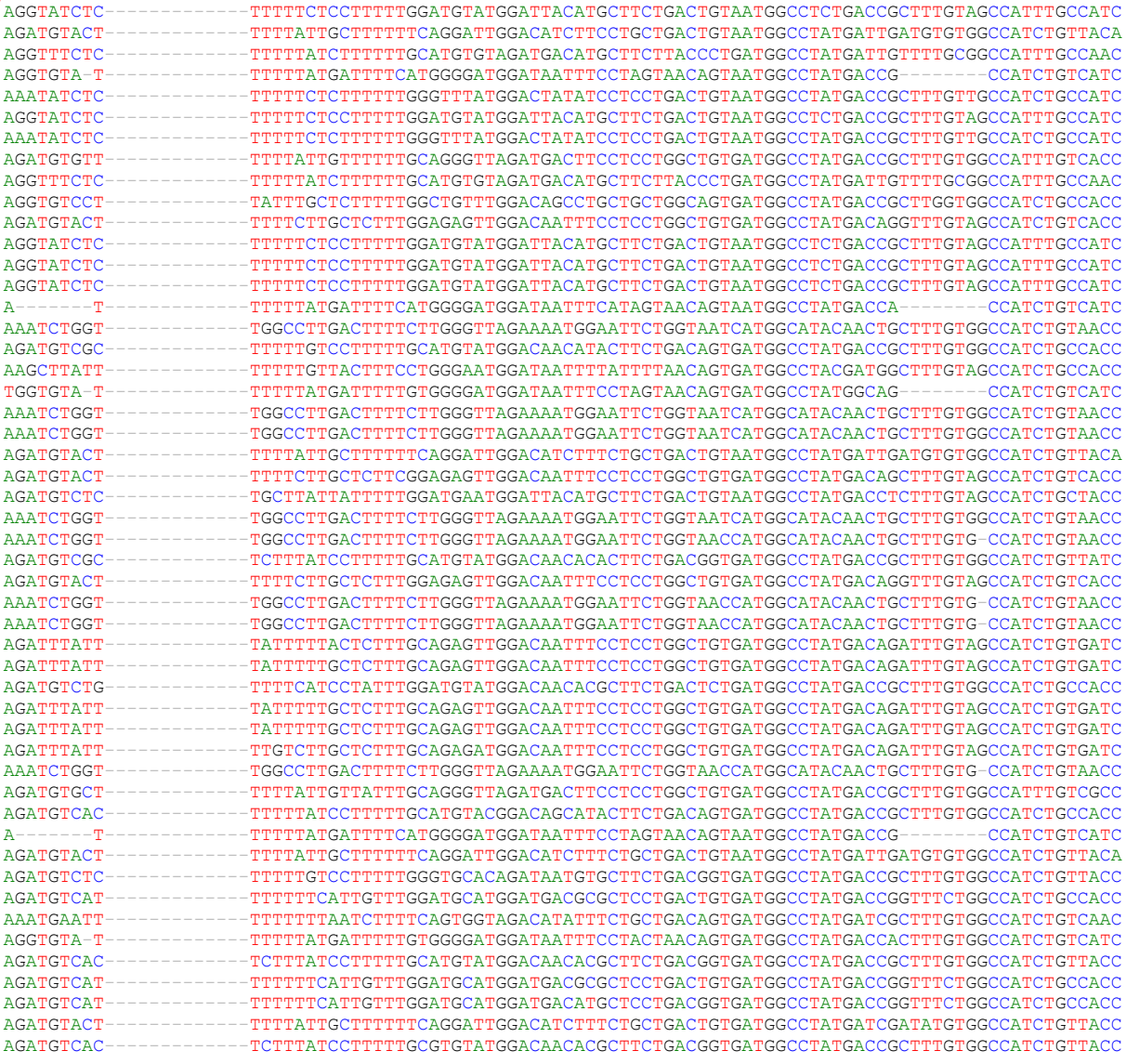


Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

↺ Back to overview

```
AAATCTGGT--------------TTGCCTTGGCTTTTCTTGGGTTAGAAAATGGAATTCTGGTAACCATGGCCTATGACCGCTTTGTGGCCATCTGTCACC
AGGTGTATT--------------TTTTTATGATTTTTGTGGGGATGGATAATTTCCTGCTGGCCGTGATGGCCTATGACCGCTTCGCGGCCATCTGCCATC
AGATTTATT--------------TTTTCTTGCTCTTTGCAGAGTTGGACAATTTCCTCCTGGCTGTGATGGCCTATGACAGGTTTATAGCCATCTGTGATC
AAGTGTATT--------------TTTTTGTACTACTTGTAATGTTGGACATCTTTCTATTGACAGTGATGTCCTATGACCGTTATGTGGCTATCTGCCACC
AGATGTCTT--------------TTTTCATCCTATTTGGATGCATGGACAACACGCTTCTGACAGTGATGGCCTATGATCGCTTTGTGGCCATCTGCCACC
AGATGTCTT--------------TTTTCATCCTATTTGGATGCATGGACAACACGCTTCTGACAGTGATGGCCTATGATCGCTTTGTGGCCATCTGCCACC
AGATGTCTT--------------TTTTCATCTTCTTTGGATGTATGGACAGCGTGCTTCTGAGTGTGATGGCCTATGACCGCTTTGTGGCCATCTGCCACC
AAATCTGGT--------------TTGCCTTGGCTTTTCTTGGGTTAGAAAATGGAATTCTGGTAACCATGGCCTATGACCGCTTTGTGGCCATCTGTCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATGGATGAATATCTACTCACTGTGATGACCTATGACAGCTTTGTGGCCATCTCTCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATTGATGAATTTCTACTCACTGTGATGACCTACGACAGTTTTGTGGTCATCTCTCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATGGATGAATTTCTACTCACTGTGATGACCTATGACAGCTTTGTGGCCATCTCTCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATTGATGAATTTCTACTCACTGTGATGACCTACGACAGTTTTGTGGCCATCTCTCACC
AGATGTACT--------------TCTTCTTGCTCTTTGGAGGGTTGGACAAGTTCCTCCTGGCTGTGATGGCCTATGACAGGTTTGTAGCCATCTGTCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
AAATGAATT--------------TTTTTTTAATCTTTTCGGTGGTAGACATATTTCTGCTGACTGTGATGGCCTATGATCGCTTTGTGGCCATCTGTCAAC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATTGATGAATTTCTACTCACTGTGATGACCTACGACAGTTTTGTGGCCATCTCTCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATTGATGAATTTCTACTCACTGTGATGACCTACGACAGTTTTGTGGCCATCTCTCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTGGCCATCTGCCACC
ATATACATT--------------TATTTTT-TTCTTCATAACACTGGACATCTTTTTGTAGACTGTGATGTCCAATGACCTCTTTGTAGCCATCTGCCACC
AGATGTCAT--------------TTTTTTTAATGTTTGCATGTATGGATGCCATGCTTCTGACTGTGATGGCCTATGACCGGTTTCTGGCCATCTGCCACC
GGGTGTATT--------------TTTTAGTAATTTTTGCTGGAATTGATGAATTTCTACTCACTGTGATGACCTACGACAGTTTTGTGGCCATCTCTCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATGGATGAATTTCTACTCACTGTGATGACCTATGACAGCTTTGTGGCCATCTCTCACC
AGGTGTATT--------------TATTTTTGGCCATTGTCACAGTGGATAATTTGCTACTGAGCATAAGGCCATGTGACTGCTTTGTGGCCATCTGCCACC
GGGTGTATT--------------TTTTAATCCTTTTTGCTGGAATGGATGATTTTCTACTCACTGTGATGGCCTATGACTGCTTTGTGGCCATCTGTCACC
GGGTGTATT--------------TTTTAATAATTTTTGCTGGAATGGATGAATTTCTACTCACTGTGATGACCTATGACAGCTTTGTGGCCATCTCTCACC
```

Figure 1 (cont.)

↩ Back to overview

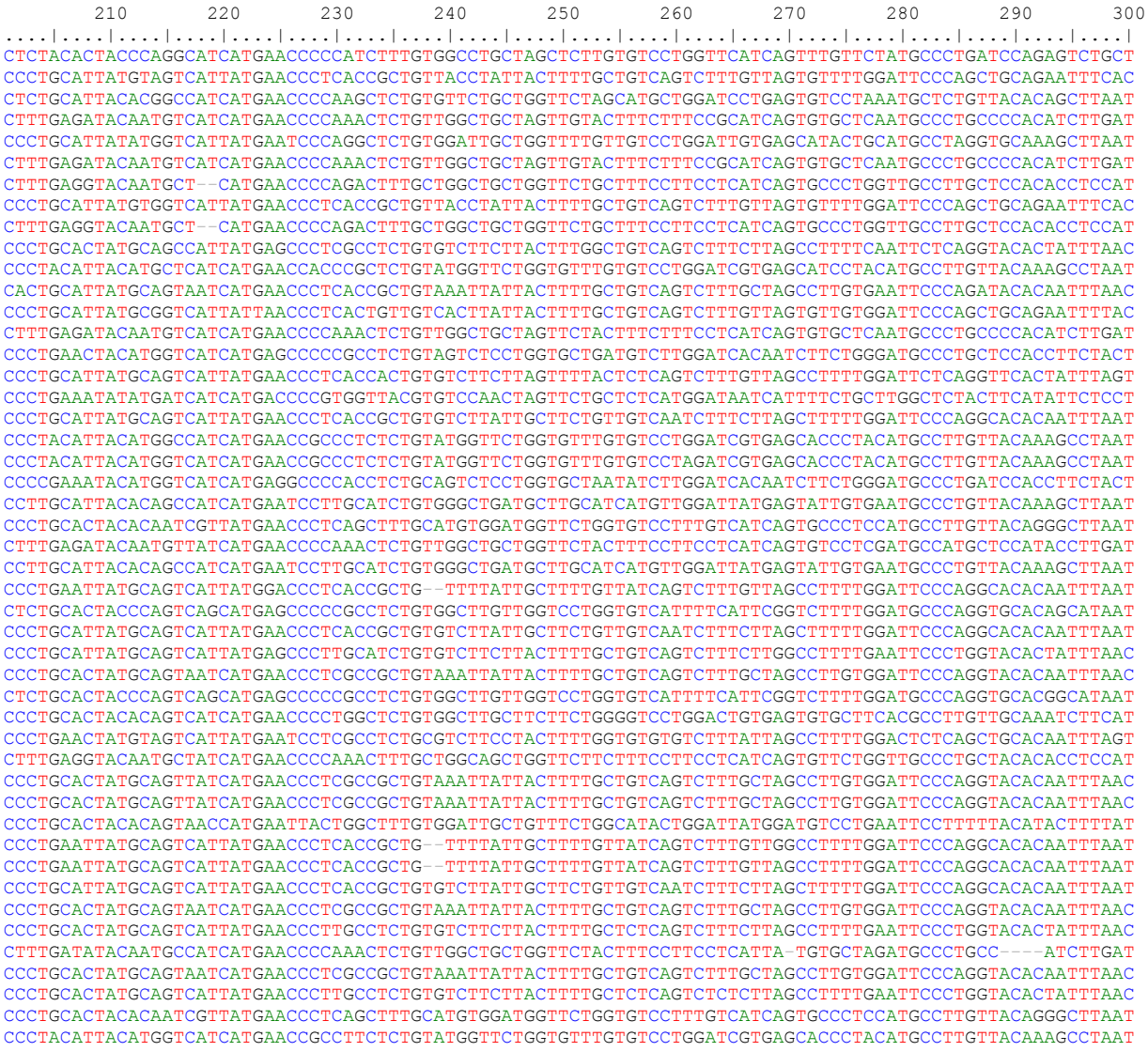


Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

⮌ Back to overview

```
CTTTGAGATACAATGTCATCATGAACCCCAAACTCTGTTGGCTGCTGGTTCTACTTTCCTTCCTCATCAGTGTCCTTGGTGCCCTGCTCCACACCTTGAT
CCCTGAGATACATGGTCATCATGACCCCTCGGCTATGTGTCCTGCTGGTTCTGGTGTCCTGGGTCATTATTTTCTGCTTCTCCTTGCTTCATATTCTCCT
CCCTGCACTACACAGTTGTCATGAACCCCTGGCTCTGTGGGTTGCTTCTTCTGGGGTCCTGGACTGTGAGTGTTCTTCATGCTTTGTTGCAAACCTTCAT
CTCTGCATTACATGGTCATCATGAATCCCCAGCTCTGTGTATTGCTGGTTCTAGCATGCTGGATCCTGAGTGTCCTAAACGCTCTGTTACACAGCTTAAT
CACTGCATTATGCAGTAATCATGAACCCTCACCGCTGTAAATTATTACTTTTGCTGTCAGTCTTTGCTAGCCTTGTGAATTCCCAGATACACAATTTAAC
CACTGCATTATGCAGTAATCATGAACCCTCACCGCTGTAAATTATTACTTTTGCTGTCAGTCTTTGCTAGCCTTGTGAATTCCCAGATACACAATTTAAC
CGCTGCATTATATGGCCATTATGAACCCTCACCGCTGTCACTTATTACTTTTGATGTCAGTCTCTGTTAGTGCTGTGGATTCCCAGATGCAGAATTTAAC
CTTTGAGATACAATGTCATCATGAACCCCAAACTCTGTTGGCTGCTGGTTCTACTTTCCTTCCTCATCAGTGTCCTTGATGCCCTGCTCCACACCTTGAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
TTCTGAACTACACAGTCATCACTAACCCTTGACACTGTAGTCTCCCGGTTCTGATATCTTGGCTTATTATTGTATGAGACTCCCTGCTCCACATTCTAAT
TTCTGAACTACACAGTCATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGACTCCCTGCTCCACATTCTAAT
CCCTGAACTACACAGTTATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGACTCCCTGCTCCACATTCTAAT
TTCTGAACTACACAGTCATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGACTCCCTGCTCCACATTCTAAT
CCCTGCATTACACAGTCATCATGAACCGCCCACTCCGTATGGTGCTGGTGTTGGTGTCCTGGGTCATGAGTGTCCTGCACGCCTTGTTACAAAGCCTGAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CCCTGCATTACATGGTCATTATGAATCCCCGGTTCTGCGGATTGCTGATTTTGGTGTCCTGGACTGTGAGCATCCTGCATGCCTTGGTGCAAAGCTTAAT
TTCTGAACTACACAGTCATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGATTCCCTGCTCCACATTCTAAT
TTCTGAACTACACAGTCATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGATTCCCTGCTCCACATTCTAAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CTCTCCATTACACGGTCATCATGAACCTCCTGCTCTCTGTGTTAATGGTTTTTGCATGCTGGATTGTGAGTGTCCTAA-TGCCCTGTTACACAGCTTAAT
CTCTGCACTATGCAATCATTATGAATCCTCGCCTCTGTGTCTTACTACTTTTGCTGTGTGTCTTTATCAGCCTTTTGGACTCTCAGCTTCACAATTTAGT
TTCTGAACTACACAGTCATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGATTCCCTGCTCCACATTCTAAT
CCCTGAACTACACAGTTATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGACTCCCTGCTCCACATTCTAAT
CCCTGAAATACATGGCCCTCATGAACTCCAATGTCTGTGTTTTCTTAGTTCTCATATCATGGCTCATCATGTTCTTGGTTTCCCTGCTTCATATTTTAAT
CCCTAAACTACAGAGTCATGATGAACTGTTGACACTCTAGTCTCCTGGCTCTCATGTCGTGGTTTATTATTGCCTGGGTCTCCTTGTTTCACATTCTACT
CCCTGAACTACACAGTTATCACTAACCCTTGACACTGTAGTCTCCTGGTTCTGATATCTTGGCTTATTATTGTATGAGACTCCCTGCTCCACATTCTAAT
```

Figure 1 (cont.)

↩ Back to overview

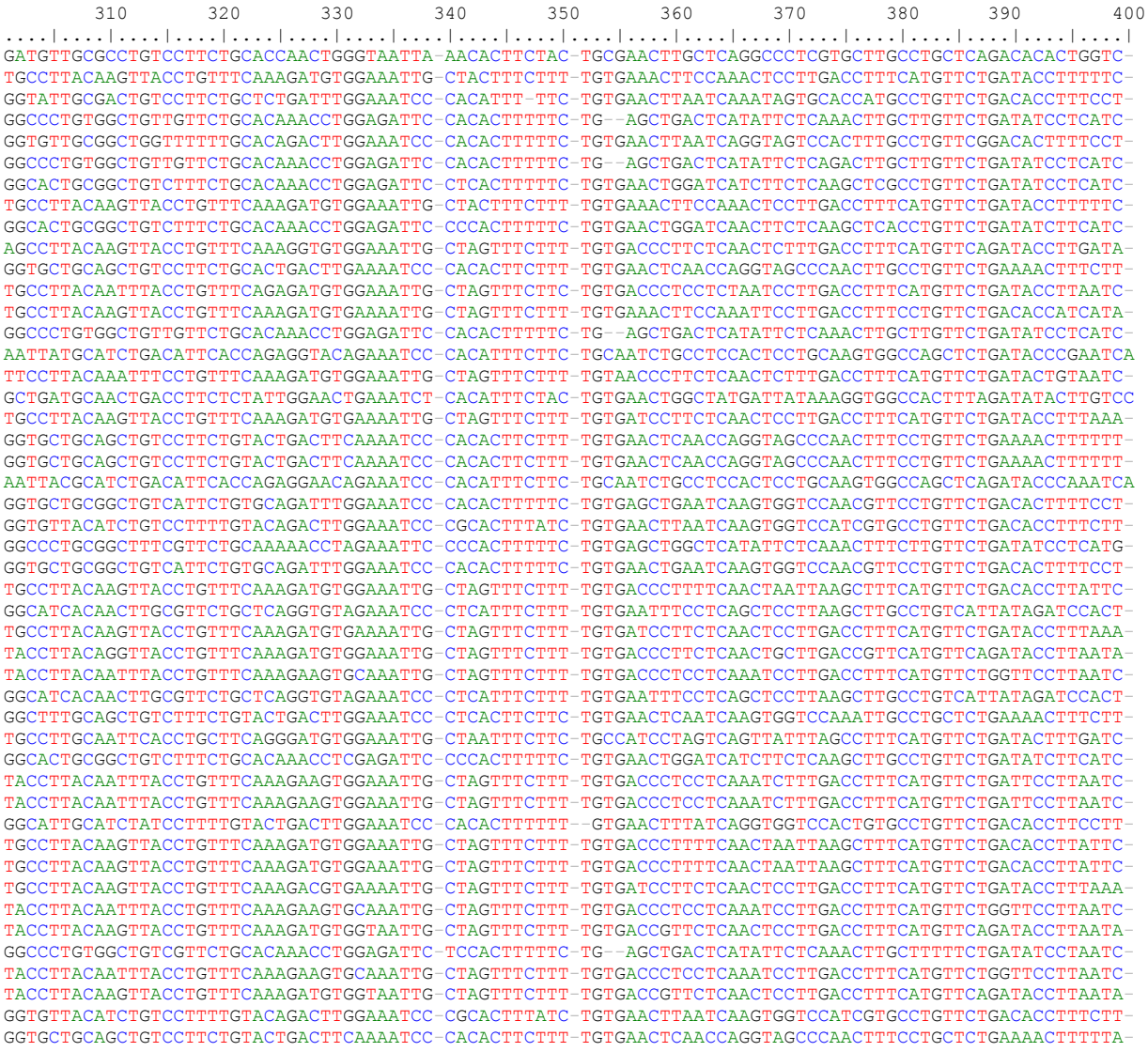


Figure 1 (cont.)

Back to overview

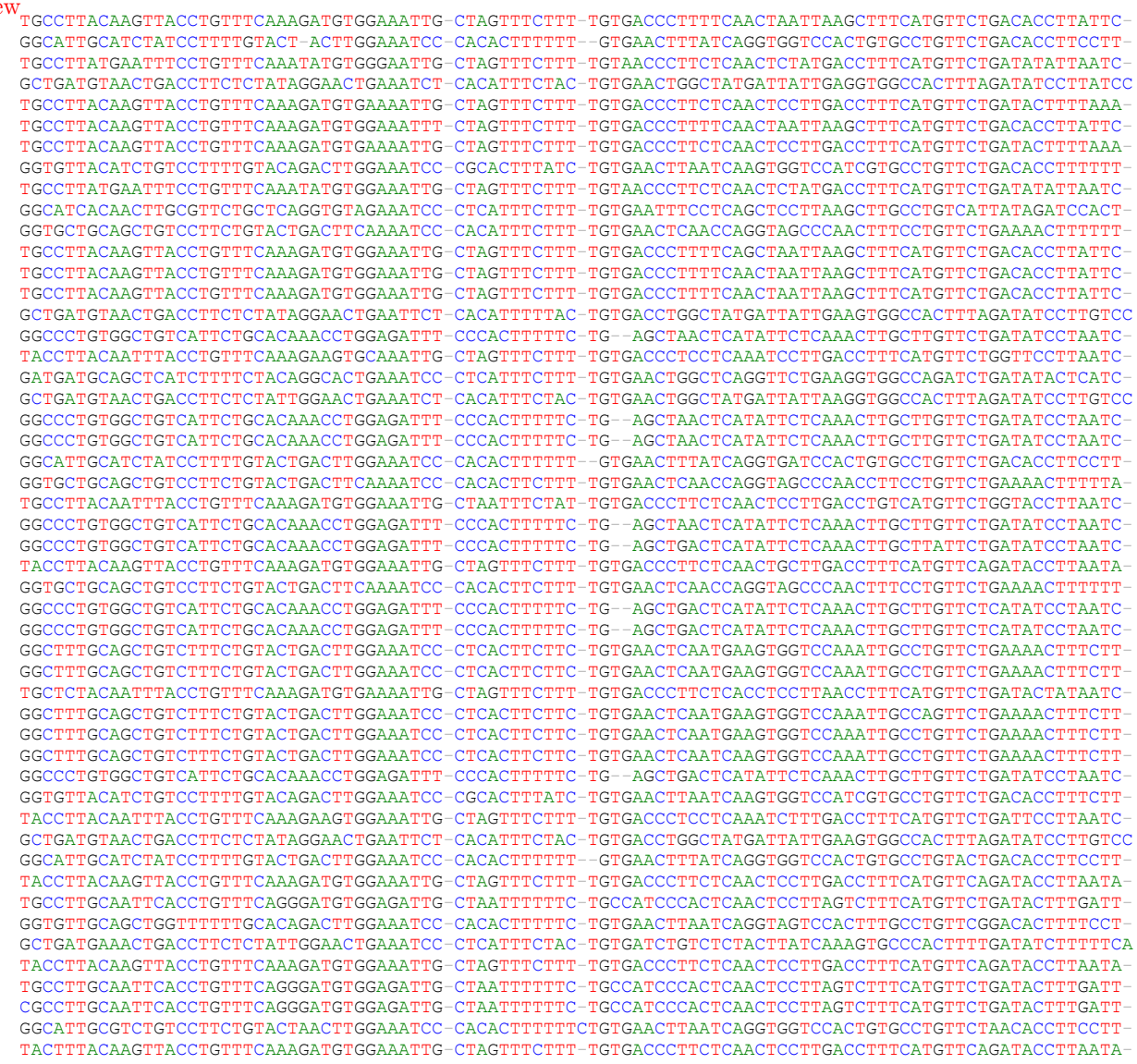


Figure 1 (cont.)

Back to overview

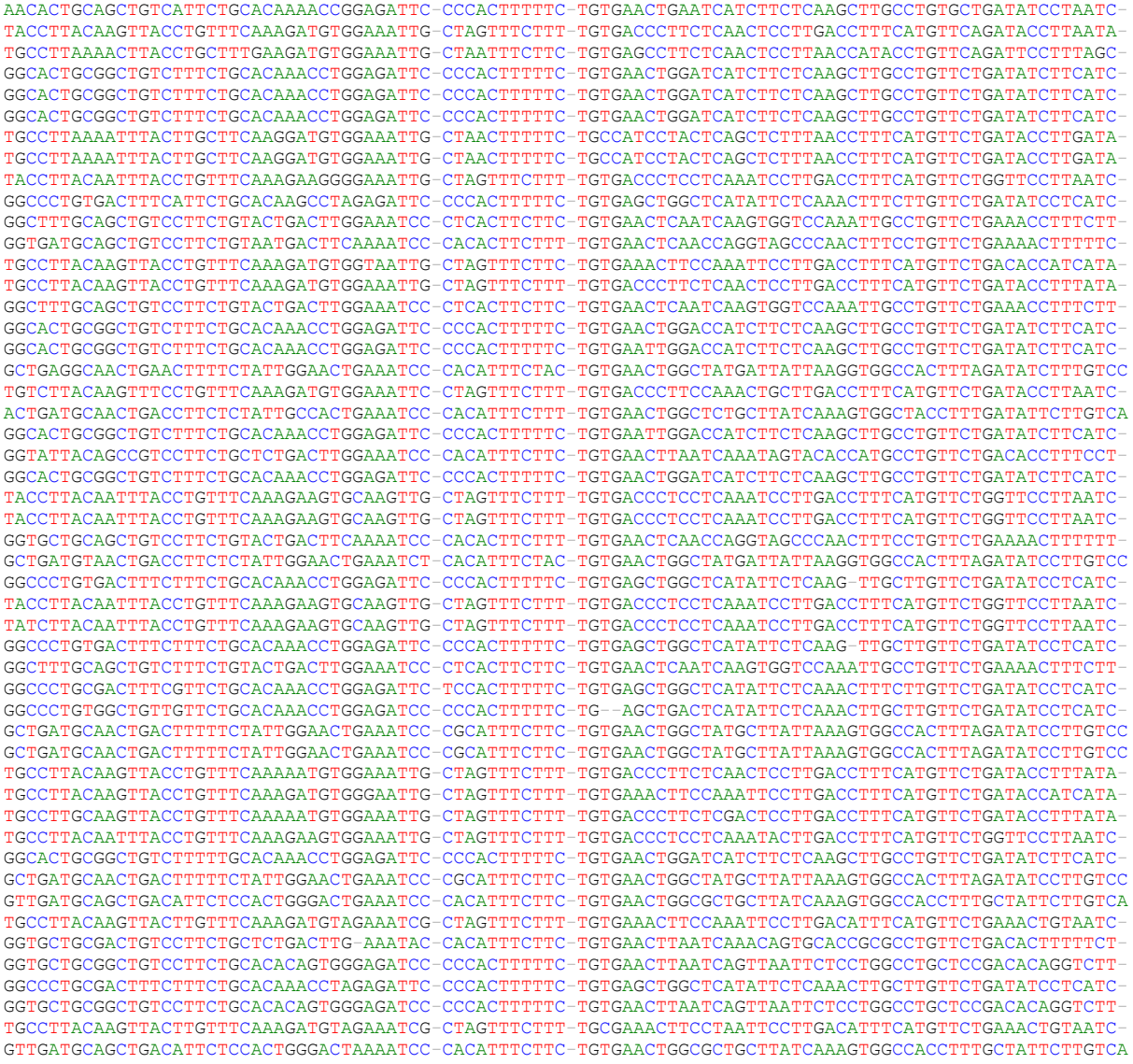


Figure 1 (cont.)

↺ Back to overview

```
GGCCCTGCGACTTTCTTTCTGCACAAACCTAGAGATTC-CCCACTTTTTC-TGTGAGCTGGCTCATATTCTCAAACTTGCTTGTTCTGATATCCTCATC-
GTTGATGCAGCTGACATTCTCCACTGGGACTGAAATCC-CACATTTCTTC-TGTGAACTGGCGCTGCTTATCAAAGTGGCCACCTTTGCTATTCTTGTCA
GGCTTTGCAGCTGTCTTTCTGTACTGACTTGGAAATCC-CTCACTTTTTC-TGTGAACTCAATCAAGTGGTCCAAATTGCCTGTTCTGAAAACTTTCTT-
GGTGCTGCGACTGTCCTTCTGCTCTGACTTGGAAATAC-CACATTTCTTC-TGTGAACTTAATCAAACAGTGCACCGTGCCTGTTCTGACACTTTTCCT-
TGCCTTACAATTTACCTGTTTCAGAGATGTGGAAATTG-CTAGTTTCTTC-TGTGACCCTCCTCTAATCCTTGACCTTTCATGTTCTGATACCTTAATC-
TGCCTTACAATTTACCTGTTTCAGAGATGTGGAAATTG-CTAGTTTCTTC-TGTGACCCTCCTCTAATCCTTGACCTTTCATGTTCTGATACCTTAATC-
TGCCTTACAAGTTACTTGTTTCAAAGATGTAGAAATCG-CTAGTTTCTTT-TGTGAAACTTCCAAATTCCTTGACATTTCATGTTCTGAAACTGTAATC-
GGCCCTGCGACTTTCTTTCTGCACAAACCTAGAGATTC-CCCACTTTTTC-TGTGAGCTGGCTCATATTCTCAAACTTGCTTGTTCTGATATCCTCATC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGAAGCCACTGAACTTTAAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTAGCACAGGTTCTCAAGGTGGCCAGCTCTGTTATCCTTATCA
GATGAAGCCACTGAACTTTAAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTAGCTCAGGTTCTCAAAGTGGCCAGCTCTGTTATCCTTATCA
GATGAAGCCACTGAACTTTCAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTACCTCAGGTTCTCAGGGTGGCCAACTCTGTTATCTTTATCA
GATGAAGCCACTGAACTTTAAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTAGCTCAGGTTCTCAAAGTGGCCAGCTCTGTTATCCTTATCA
GGTGCTGAGGCTGTCCTTCTGTACTGACTTGAAAATCC-CACACTTCTTT-TGTGAACTCAATCAGGTAGCCCAACTTGCCTGTTCTGAAAACTTTTTA-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GGTGTTGTGGCTGGTCTTTTGCATAGACTTGGAAATCC-CACATTTTTTC-TGTGAACTTAATCAGGTAGTCCACTGTGCCTGTTCAGACACTTTTCCT-
GATGAAGCCACTGAACTTTAAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTAGCTCAGGTTCTCAAAGTGGCCAGCTCTGTTATCCTTATCA
GATGAAGCCACTGAACTTTAAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTAGCTCAGGTTCTCAAAGTGGCCAGCTCTGTTATCCTTATCA
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-TGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
GATGCTGCAGCTGACCTTCTGCACTGACCTGGAAATCC-CACATTTCTTC-CGTGAACTTAATCAGGTGATCCACTGTGTTTGCTCTGACACATTATCC-
TGCCTTGCAATTTACCTGCTTCAAGGATGTGGAAATTG-CTAATTTCTTC-TGCCATCCTAGTCAACTCCTTAGACTTTCATGTTATGATACTTTGATC-
GATGAAGCCACTGAACTTTAAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTAGCTCAGGTTCTCAAAGTGGCCAGCTCTGTTATCCTTATCA
GATGAAGCCACTGAACTTTCAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTACCTCAGGTTCTCAGGGTGGCCAACTCTGTTATCTTTATCA
AATGAAGCAATTGAATTTCTCCATAAACACCAAAATCCTCCCATTTCTTC-TGTGAACTGGCTCAGGTTCTCAAGGTGGCCAGGTCTGATACCATTGCCA
AATGAAGCCATTGACCTTCCACATGGGCACTGAAATCC-CTCATTTCTTC-TGTGAATTATCTCAGGTTCTCAAGGCAGCCAGCTCTGATACCCTTATCA
GATGAAGCCACTGAACTTTCAAATAGGCACTGAAATCC-CTCATTTTGTC-TGTGAATTACCTCAGGTTCTCAGGGTGGCCAACTCTGTTATCTTTATCA
```

Figure 1 (cont.)

Back to overview

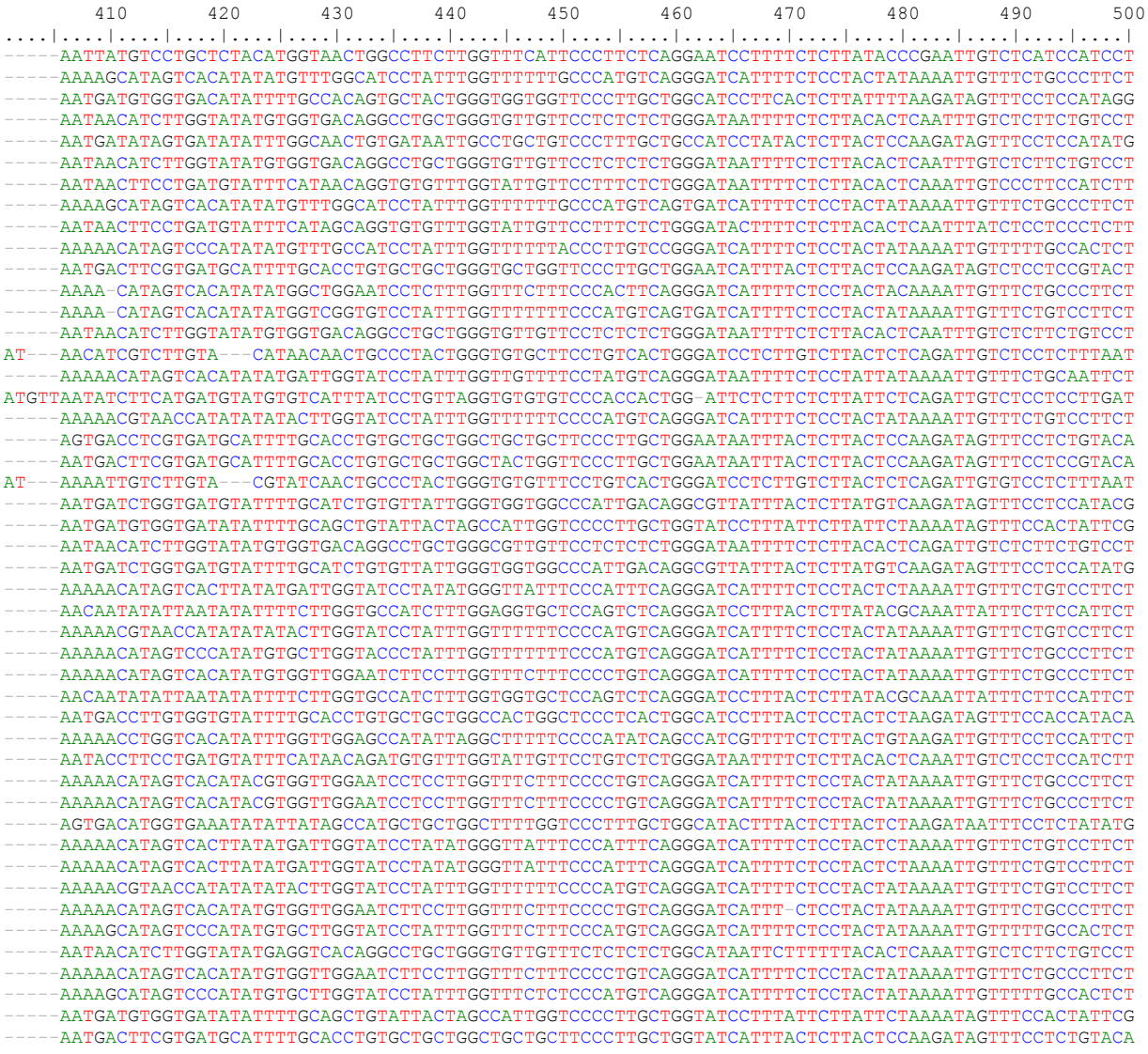


Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

↩ Back to overview

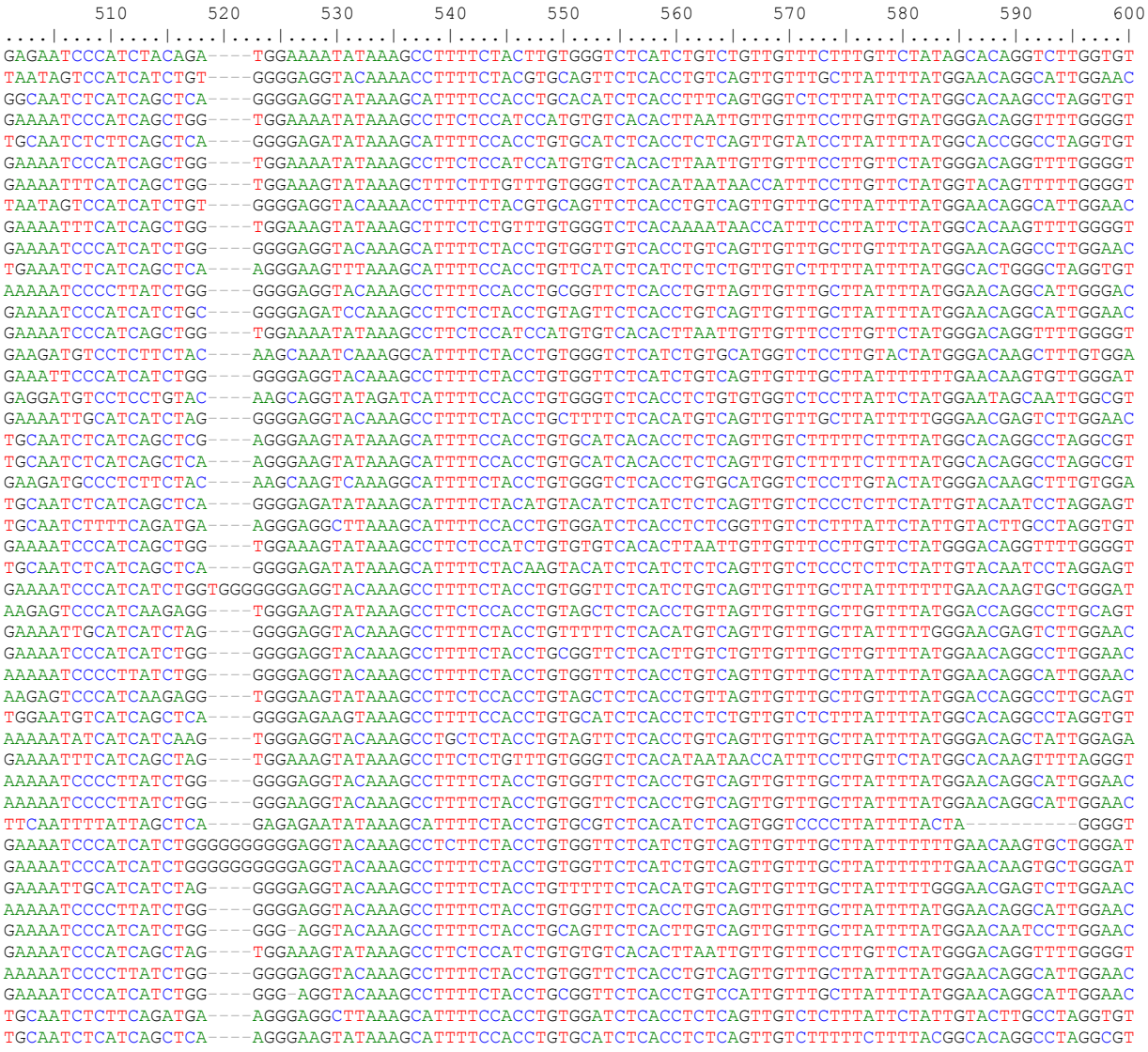


Figure 1 (cont.)

Back to overview

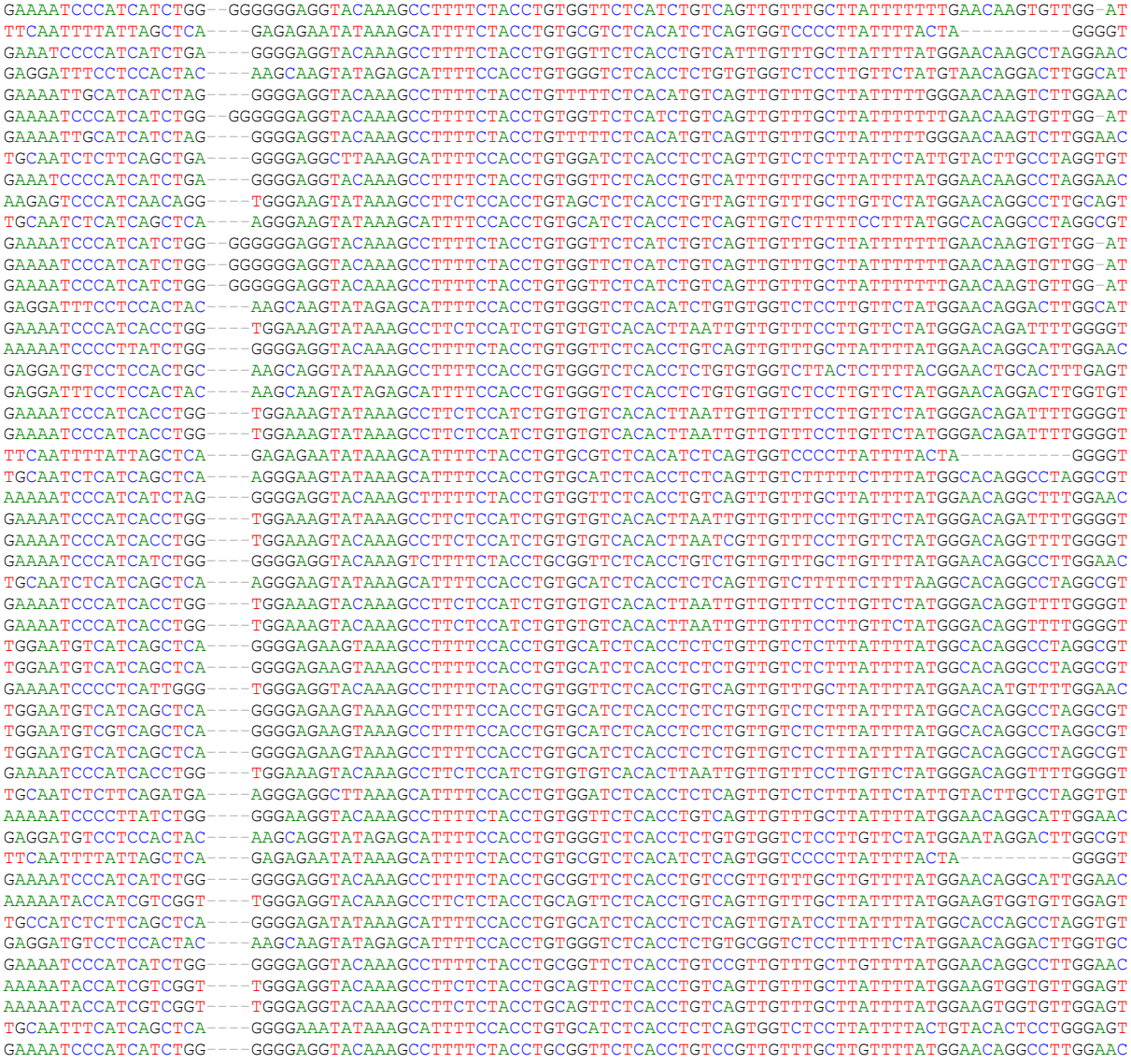


Figure 1 (cont.)

Back to overview

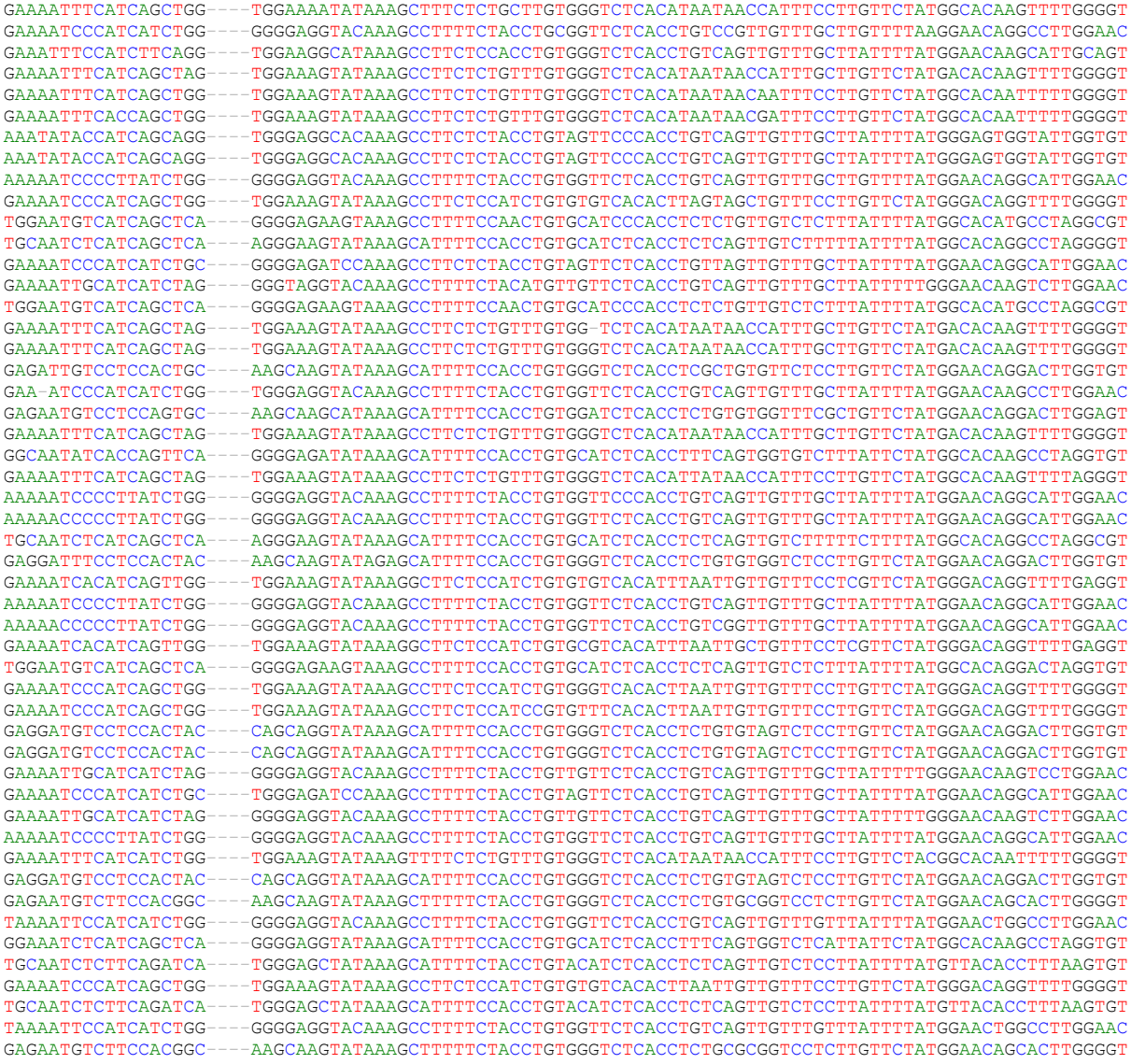


Figure 1 (cont.)

⮌ Back to overview

```
GAAAATCCCATCAGCTGG----TGGAAAGTATAAAGCCTTCTCCATCTGTGTGTCACACTTAATTGTTGTTTCCTTGTTCTATGGGACAGGTTTTGGGGT
GAGAATGTCTTCCACGGC----AAGCAAGTATAAAGCTTTTTCTACCTGTGGGTCTCACCTCTGTGCGGTCCTCTTGTTCTATGGAACAGCACTTGGGGT
TGGAATGTCATCAGCTCA----GGGGAGAAGTAAAGCCTTTTCCACCTGTGCATCTCACCTCTCTGTTGTCTCTTTATTTTATGGCACAGGCCTAGGTGT
GGAAATCTCATCAGCTCA----GGGGAGGTATAAAGCATTTTCCACCTGTGCATCTCACCTTTCAGTGGTCTCATTATTCTATGGCACAAGCCTAGGTGT
AAAAATCCCCTTATCTGG----GGGGAGGTACAAAGCCTTTTCCACCTGTGGTTCTCACCTGTTAGTTGTTTGCTTATTTTATGGAACAGGCATTGGAAC
AAAAATCCCCTTATCTGG----GGGGAGGTACAAAGCCTTTTCCACCTGCGGTTCTCACCTGTTAGTTGTTTGCTTATTTTATGGAACAGGCATTGGAAC
TAAAATTCCATCATCTGG----GGGGAGGTACAAAGCCTTTTCTACCTGTGGTTCTCACCTGTCAGTTGTTTGTTTATTTTATGGAACTGGCCTTGGAAC
GAAAATCCCATCAGCTGG----TGGAAAGTATAAAGCCTTCTCCATCTGTGTGTCACACTTAATTGTTGTTTCCTTGTTCTATGGGACAGGTTTTGGGGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
GAGAATGTGCTGGTCCTC----AAGCACGTATAAAGCATTCTCCACCTGTGGGACACACCTCTGTGTGGTGTCCTTGTTCTTTGAAACAACACTCG---T
GAGAATGTGCTGGTCCTC----AAGCAGGTATAAAGCATTCTCCACCTGTGGGTCACACCTCTGTGTGGTGTCCTTGTTCTTTGAAACA---CTTA---T
GAGAATGTGCTGGTCCTC----AAGCAGGTATAAAGTATTCTCCACCTGTGGGTCACACCTCTGTGTGGTGTCCTTGTTCTTTGAAACAACACTTG---T
GAGAATGTGCTGGTCCTC----AAGCAGGTATAAAGCATTCTCCACCTGTGGGTCACACCTCTGTGGGGTGTCCTTGTTCTTTGAAACAACACTTG---T
TGCAGTCTCATCAGCTCA----AGGGAAGCATAAAGCCTTTTCCACCTGTGCGTCTCACCTCTCAGGTGTCTCCTTATTTTATGGCACTGGCCGAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTCTTCAGCTCA----GGGGAGATATAAAGCATTTTCCACCTGTGTCTCTCACCTCTCAGTTGTATCCTTATTTTATGGCACAAGCCTAGGTGT
GAGAATGTACTGGTCCTC----AAGCAGGTATAAAGCATTCTCCACCTGTGGGTCACACCTCTGTGGGGTGTCCTTGTTCTTTGAAACAACACTTG---T
GAGAATGTGCTGGCCCTC----AAGCAGGTATAAAGCATTCTCCACCTGTGGGTCACACCTCTGTGTGGTGTCCTTGTTCTTTGAAACAACACTTG---T
TGCAATCTAATCAACTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
TGCAATCTAATCAGCTCA----GGGGAGATATAAAGCATTTTCAACCTGTTCATCTCACCTCTCAGTGATCTCCTTATTTTATGCCACTGAGCTAGGTGT
AAAAATATCATCATCAGG----TGGGAGGTGCAAAGCCTTCTCTACCTGCAGTTCTCATCTGTCAGTTGTTTGCTTATTTTATGGGACAGCTATTGGAGT
GAGAATGTACTGGTCCTC----AAGCAGGTATAAAGCATTCTCCACCTGTGGGTCACACCTCTGTGGGGTGTCCTTGTTCTTTGAAACAACACTTG---T
GAGAGTGTGCTGGTCCTC----AAGCAGGTATAAAGTATTCTCCACCTGTGGGTCACACCTCTGTGTGGTGTCCTTGTTCTTTGAAACAACACTTG---T
GAGGATATCCTCCTCTGC----AAGCAGATGTAAAGCATTTTCCACCTGTGGGTCTCACCTCTGTGTGGTCTCCTTGTTCTATGGAACAGGATTTGGGGA
GAGAATATGCTGGTCCTC----AAGCAGGTAGAAAGCATTCTCCCCCTGTGTGTTTCACCTCTGTGTGGTGTCCTTGTCTTATGGAATAGCACTTGGGGT
GAGAATGTGCTGGTCCTC----AAGCAGGTATAAAGTATTCTCCACCTGTGGGTCACACCTCTGTGTGGTGTCCTTGTTCTTTGAAACAACACTTG---T
```

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

Figure 1 (cont.)

Back to overview

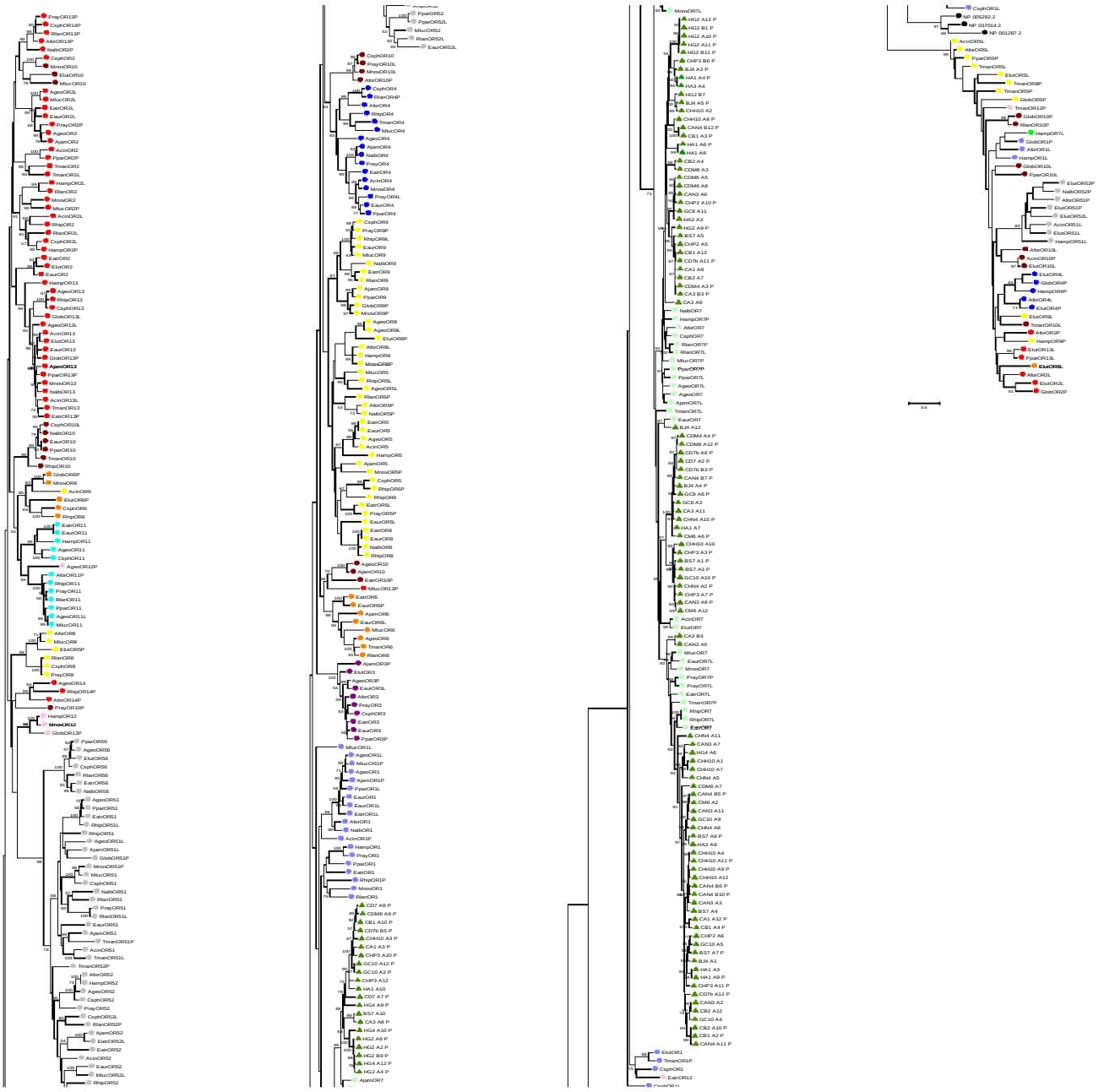


Figure 2: Maximum likelihood tree of Mammalian OR genes Maximum Likelihood tree obtained with Tamura-Nei substitution model and 1000 bootstrap replicates using representative sequences of all OR Families from the available Mammalian database (Hayden et al. 2010 and nomenclature therein), together with the Bathyergidae OR genes characterised in this study (for gene-name abbreviations see Appendix I.2). Rhodopsin-like non-OR GPCRs are used to root the tree (accession numbers NP_001287.2, NP_005292.2, NP_037014.2). Each OR family has a distinctive colour; bootstrap support values higher that 50 are indicated on corresponding branches. The tree is represented as a rectangular cladogram to allow for visualisation of internal nodes; for a global view of the tree topology please refer to Figure 2.7 in the main text of this thesis.

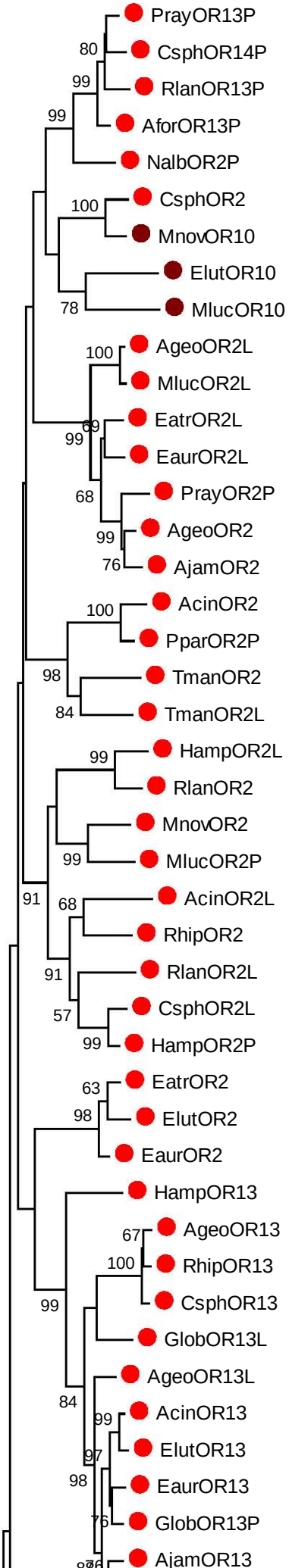


Figure 2 (cont.)

Back to overview

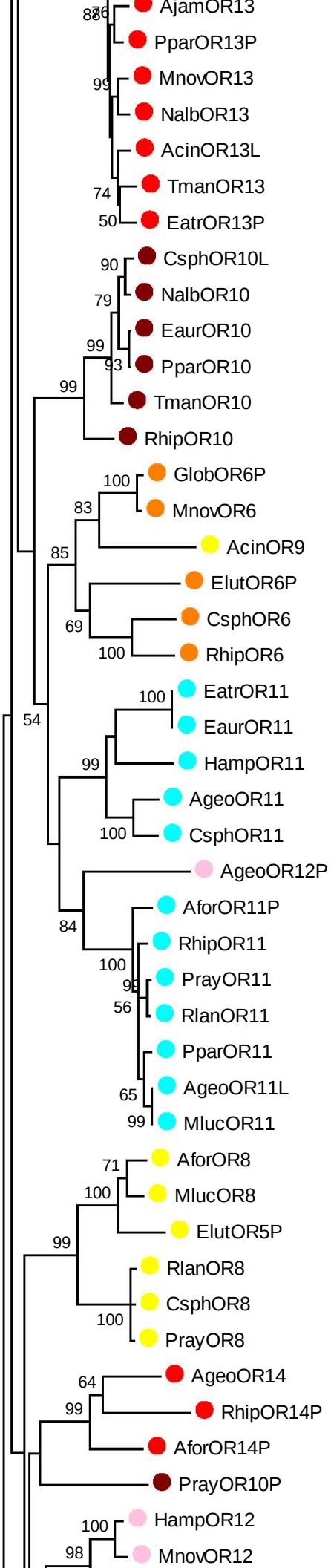


Figure 2 (cont.)

Back to overview

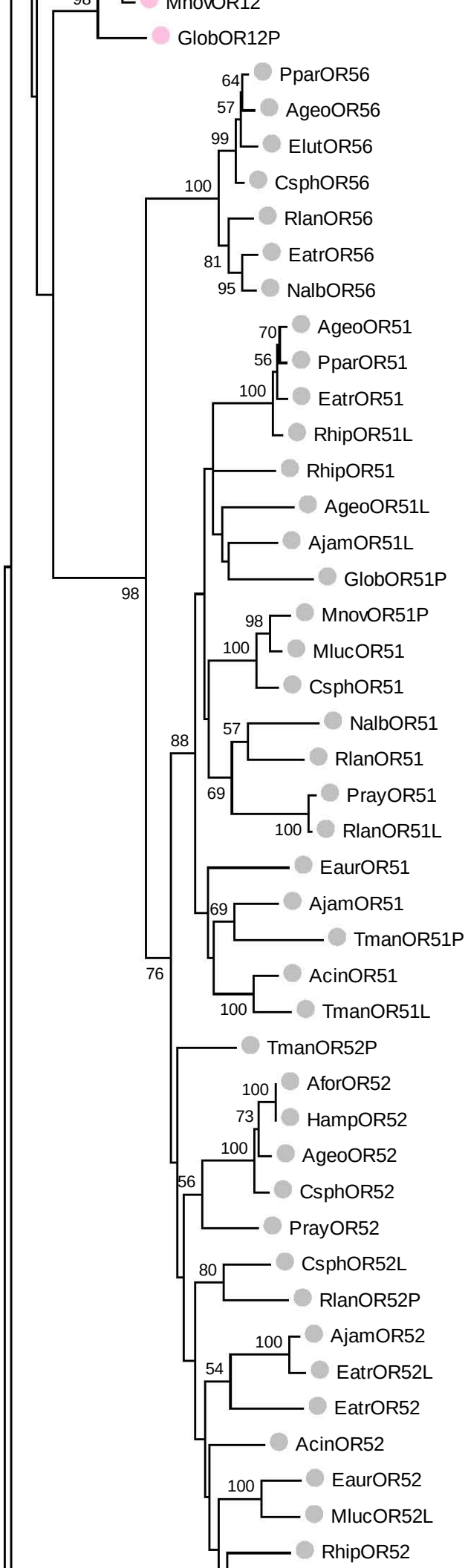


Figure 2 (cont.)

Back to overview

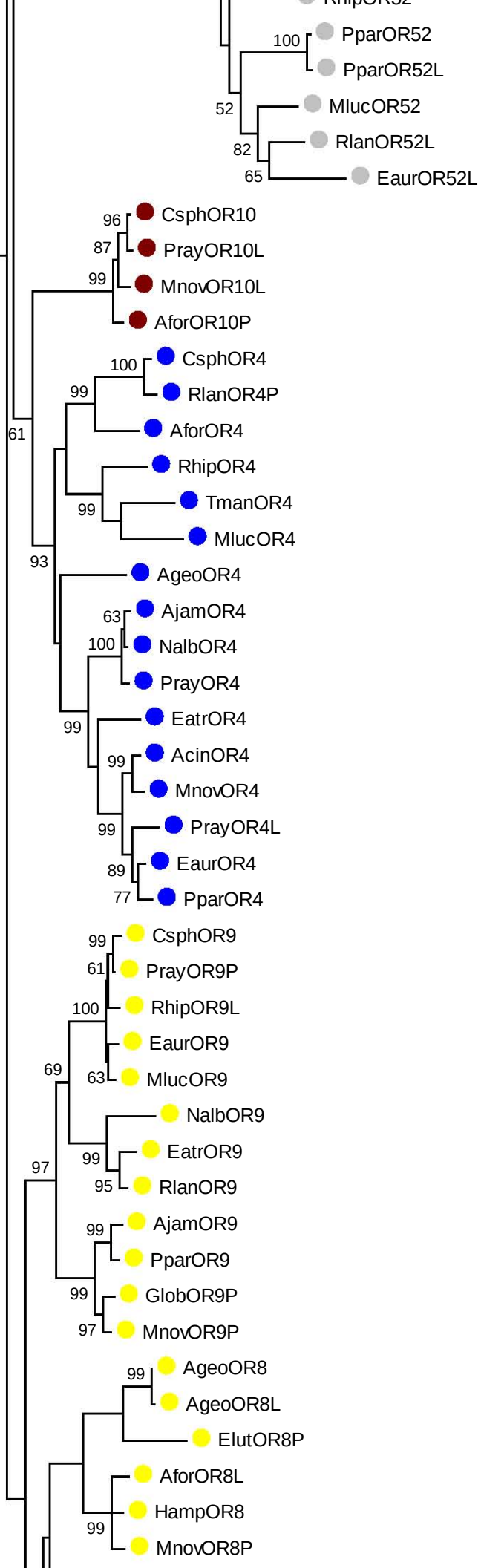


Figure 2 (cont.)

Back to overview

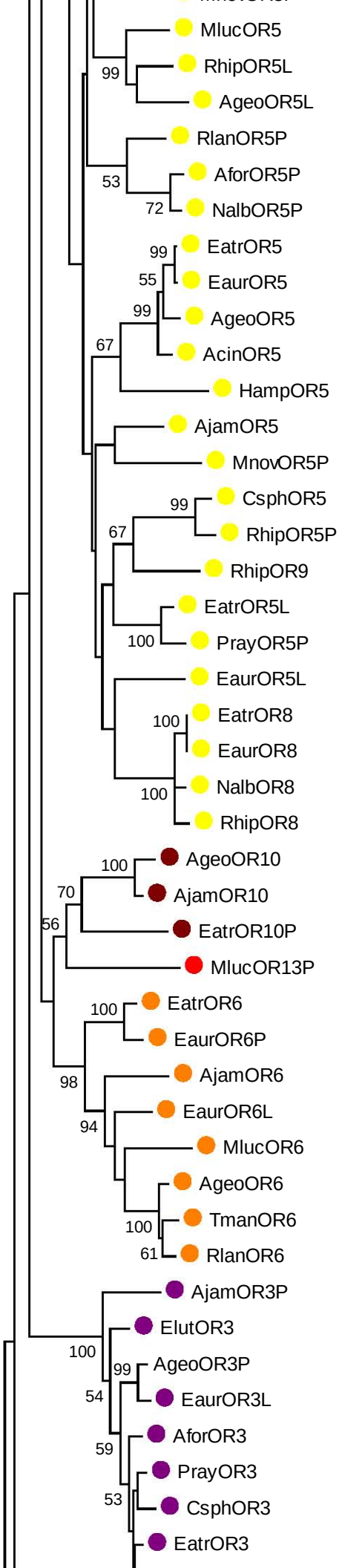


Figure 2 (cont.)

Back to overview

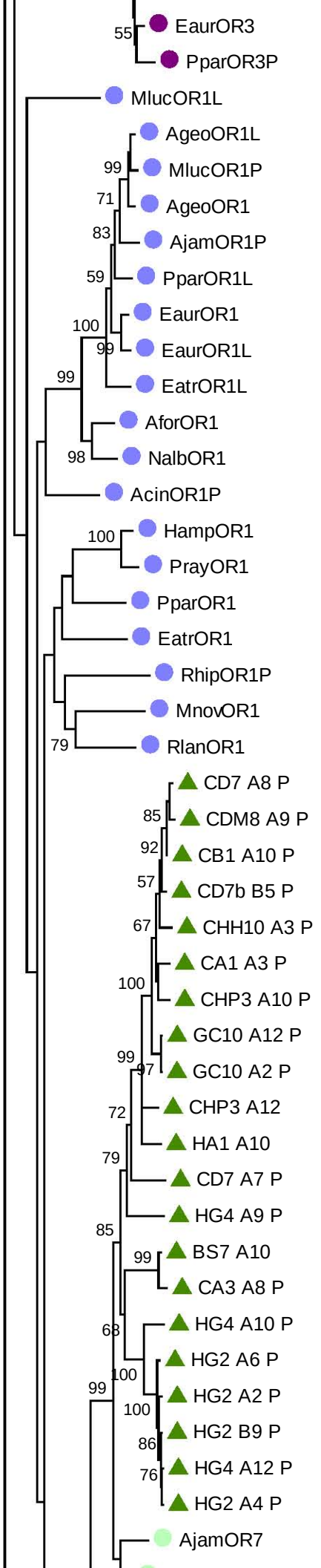


Figure 2 (cont.)

Back to overview

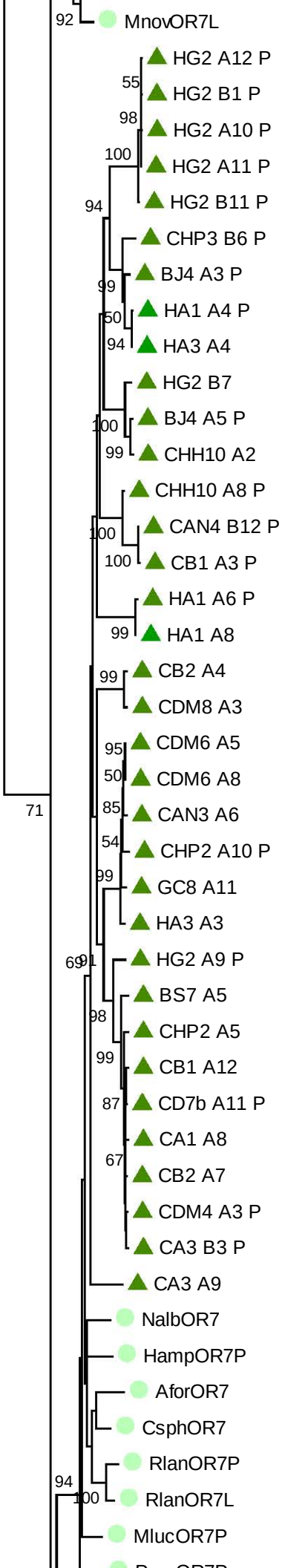


Figure 2 (cont.)

Back to overview

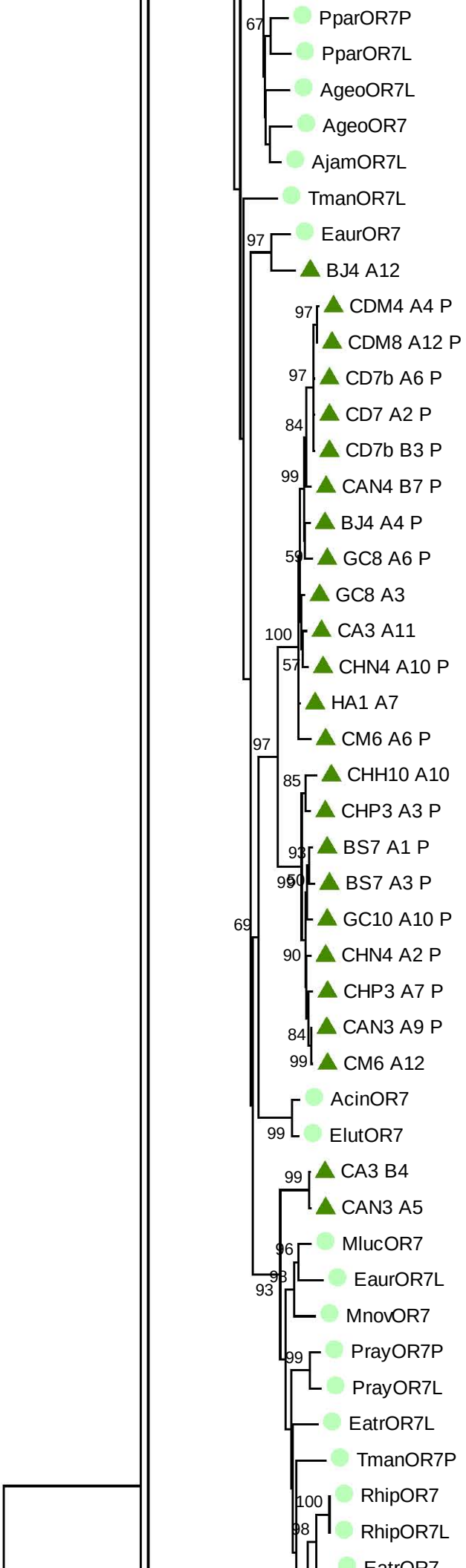


Figure 2 (cont.)

Back to overview

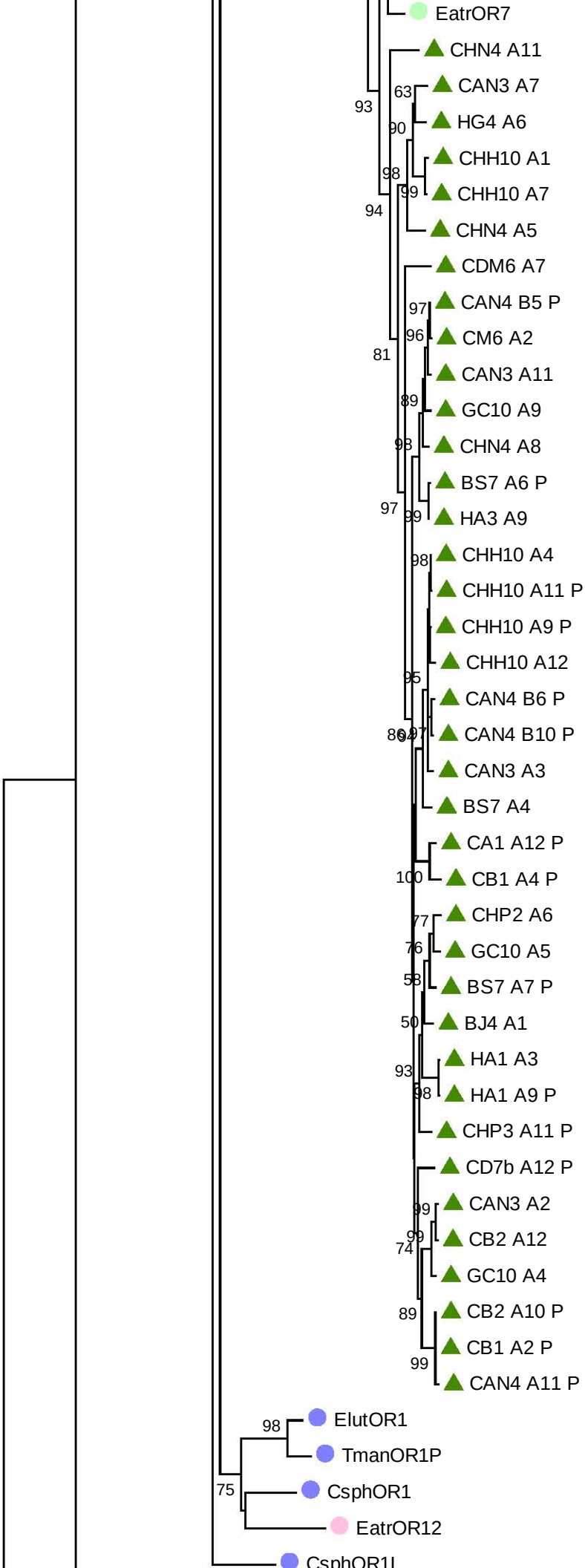


Figure 2 (cont.)

Back to overview

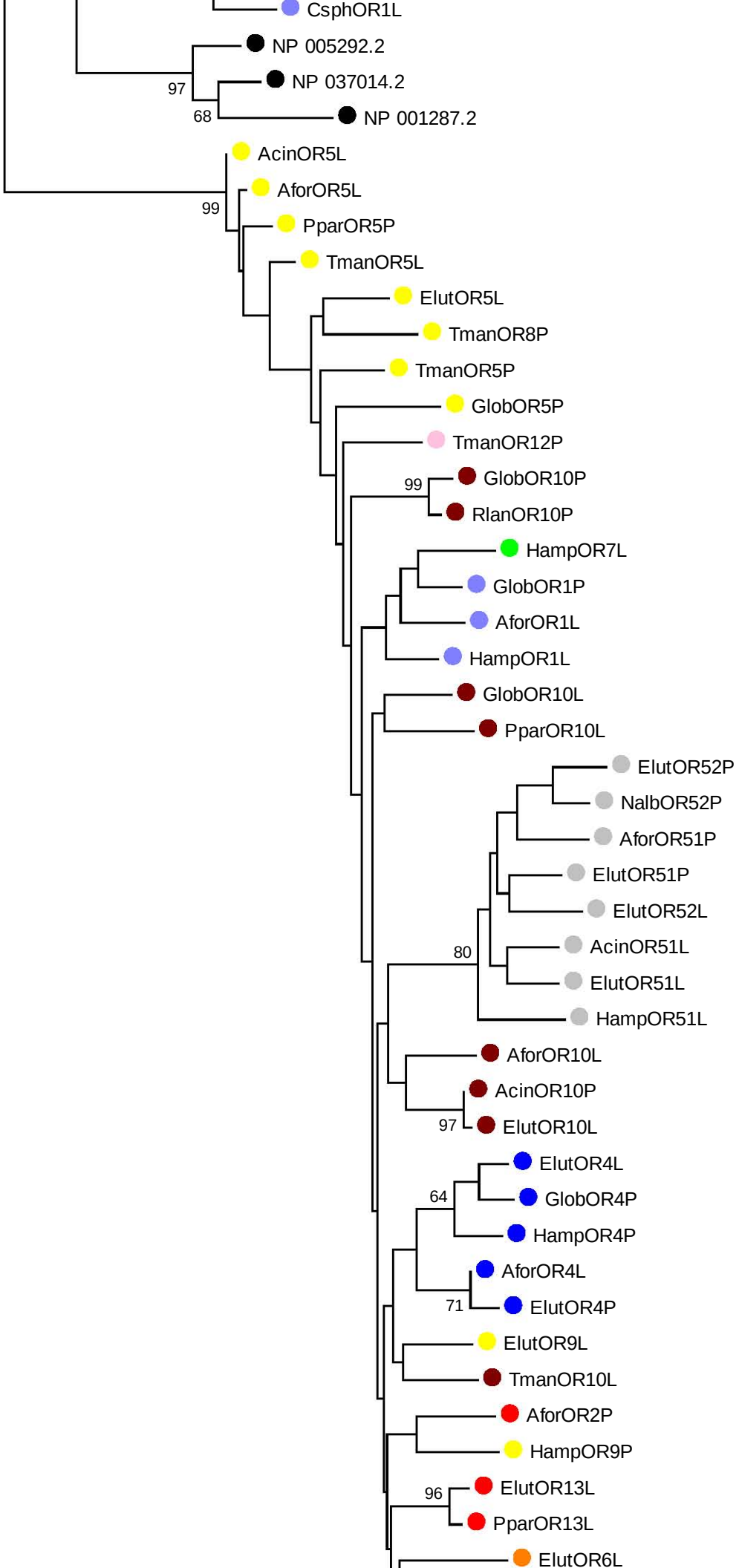


Figure 2 (cont.)

Back to overview

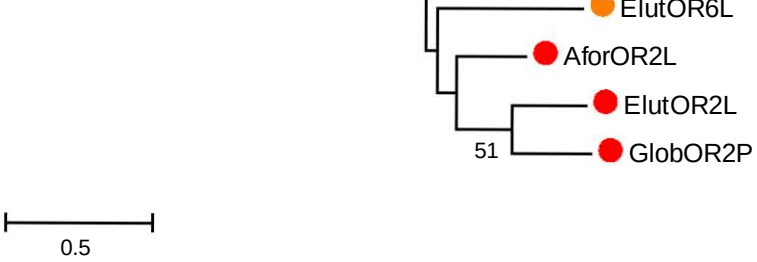


Figure 2 (cont.)
↩ Back to overview

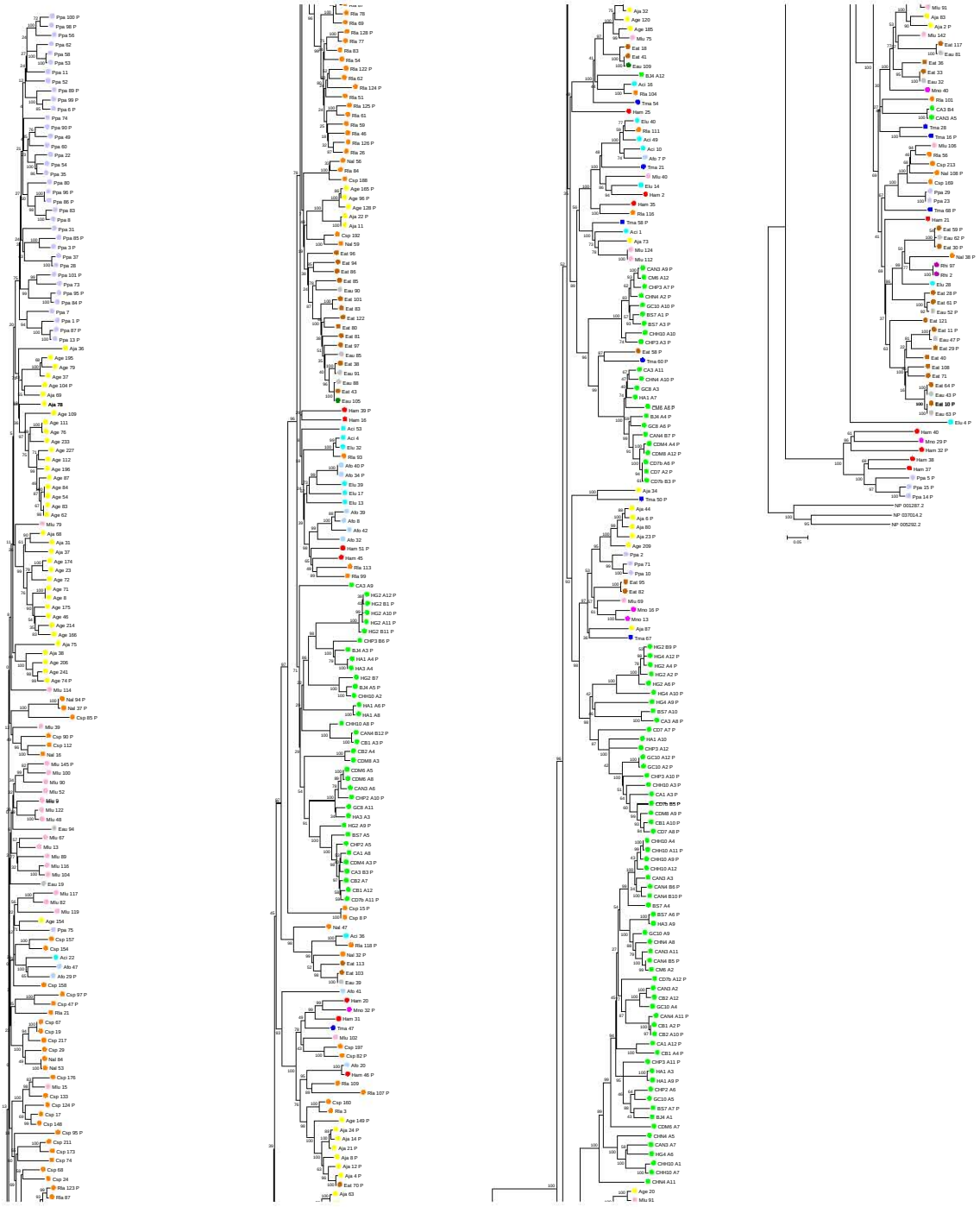


Figure 3: Neighbour-joining tree of the mammalian OR7 gene family Neighbour-joining tree (p-dist, 1000 boostrap) constructed with all known mammalian OR genes (Hayden et al. 2010). ORs from different taxonomic families have distinctive colours. Gene names are abbreviated as per Appendix I.2 and Hayden et al. 2010; rhodopsin-like non-OR GPCRs are used to root the tree (accession numbers NP_001287.2, NP_005292.2, NP_037014.2). The tree is represented as a rectangular cladogram to allow for visualisation of internal nodes; for a global view of the tree topology please refer to Figure 2.8 in the main text of this thesis.

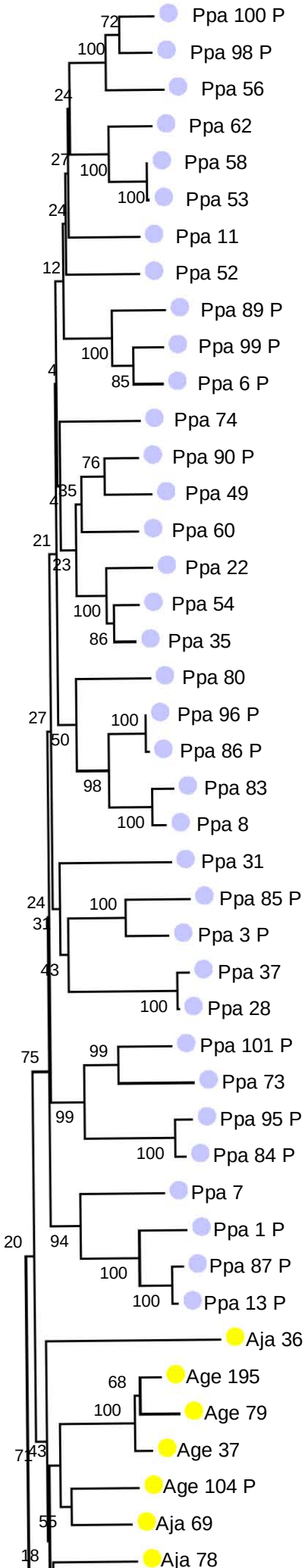


Figure 3 (cont.)

Back to overview

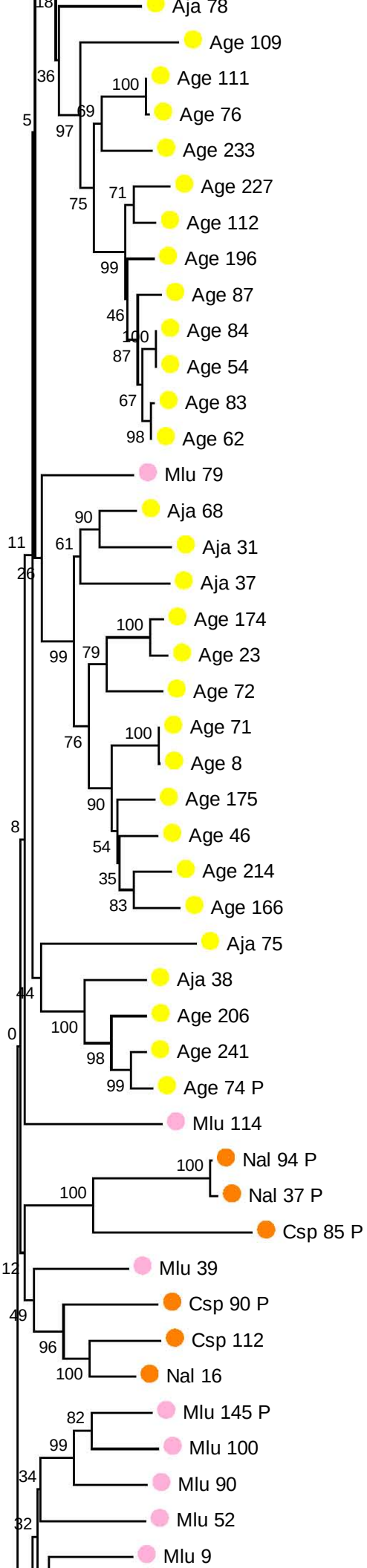


Figure 3 (cont.)

Back to overview

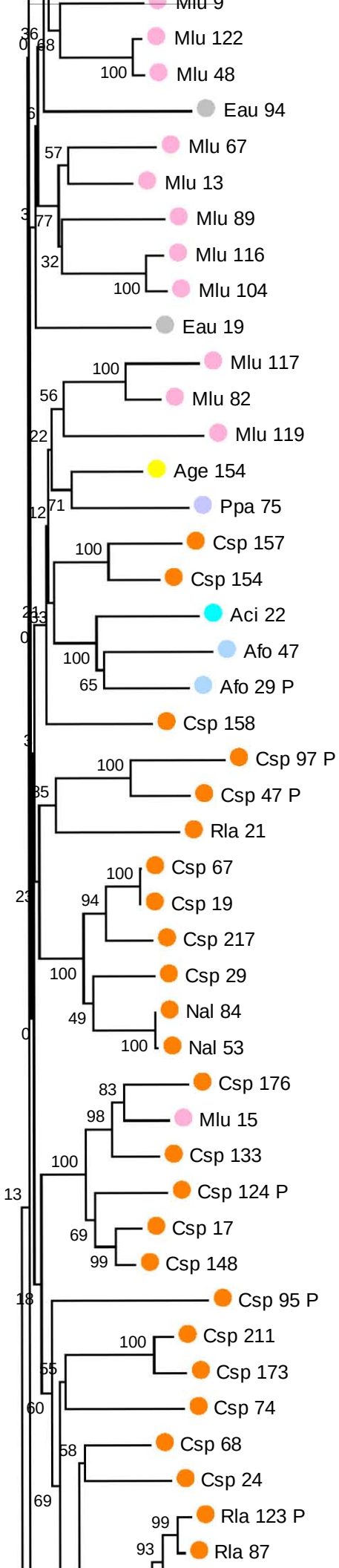


Figure 3 (cont.)

Back to overview

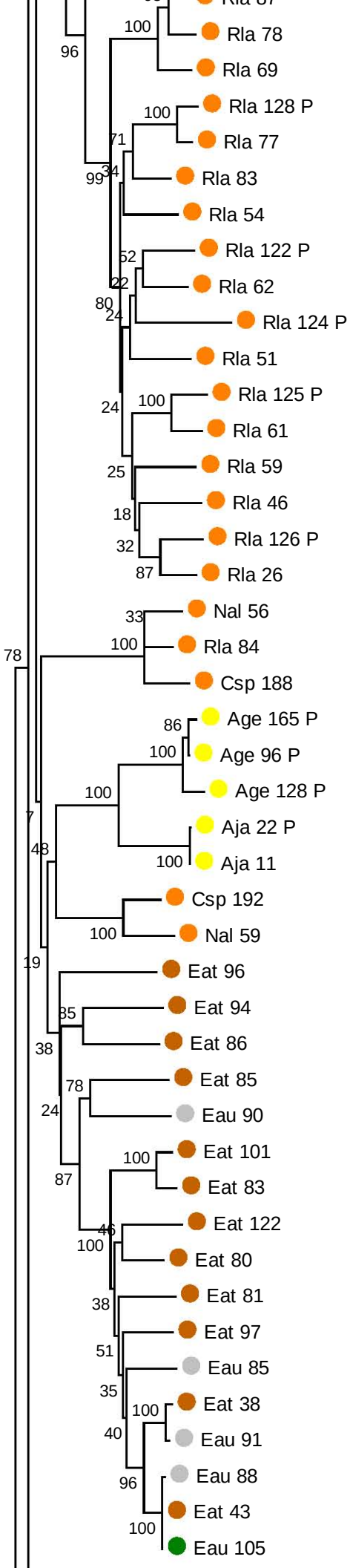


Figure 3 (cont.)

Back to overview

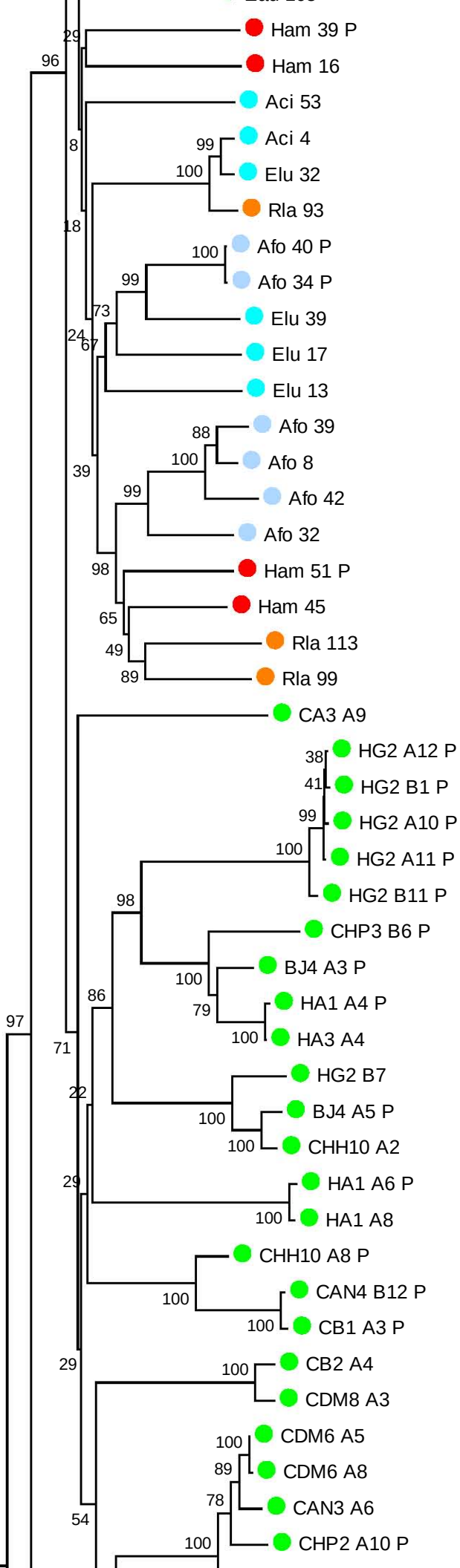


Figure 3 (cont.)

Back to overview

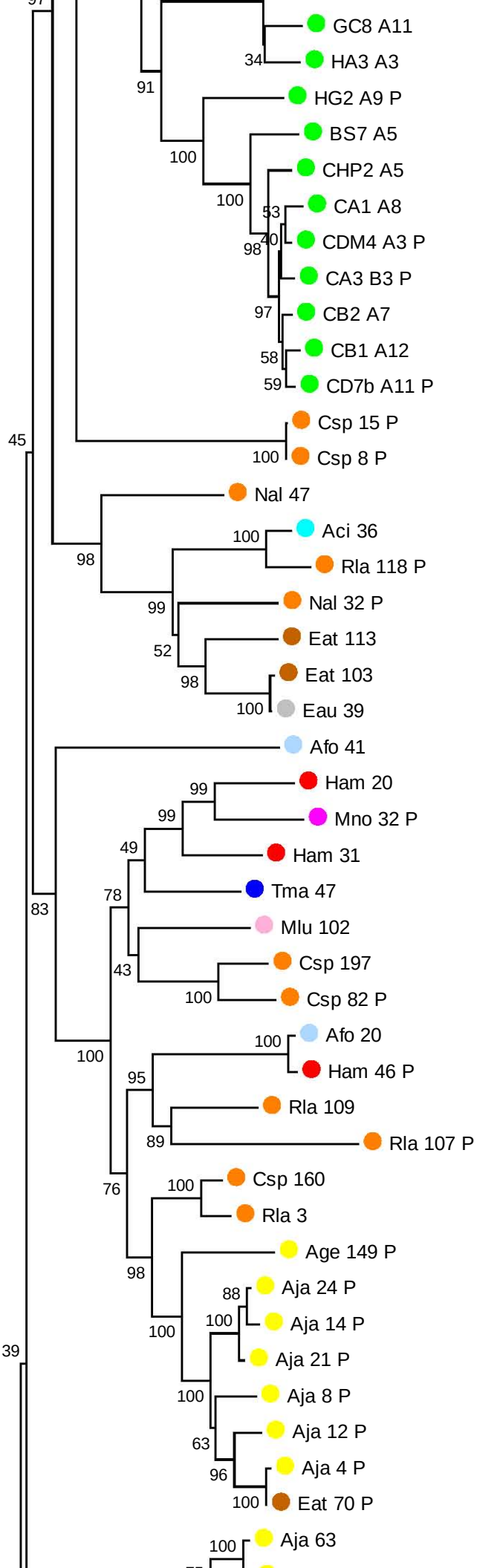


Figure 3 (cont.)

Back to overview

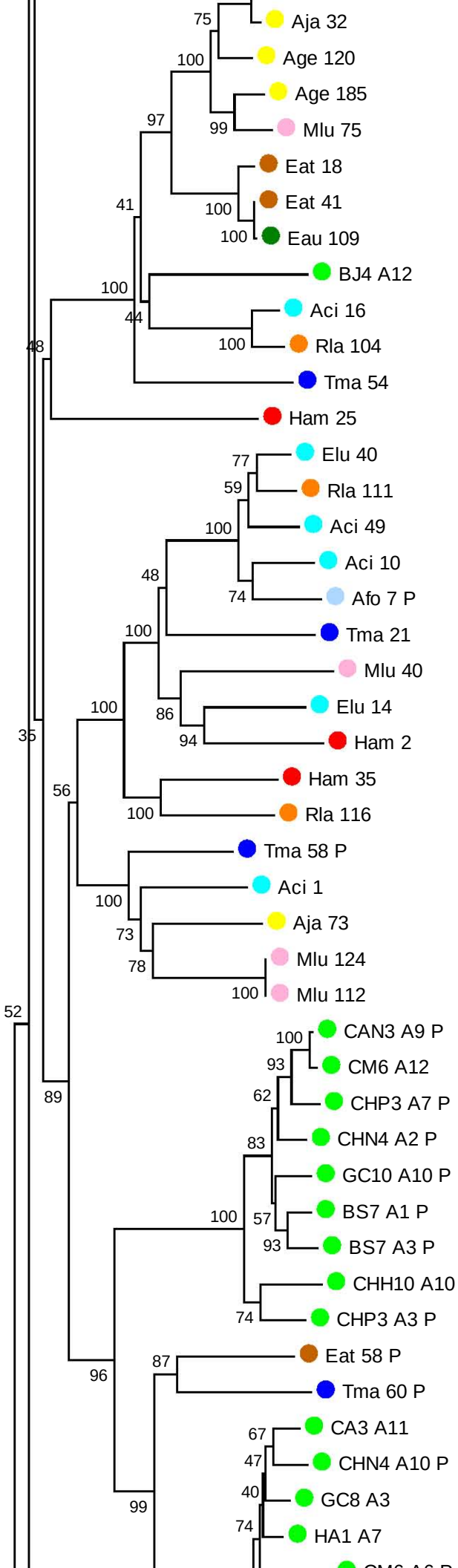


Figure 3 (cont.)

Back to overview

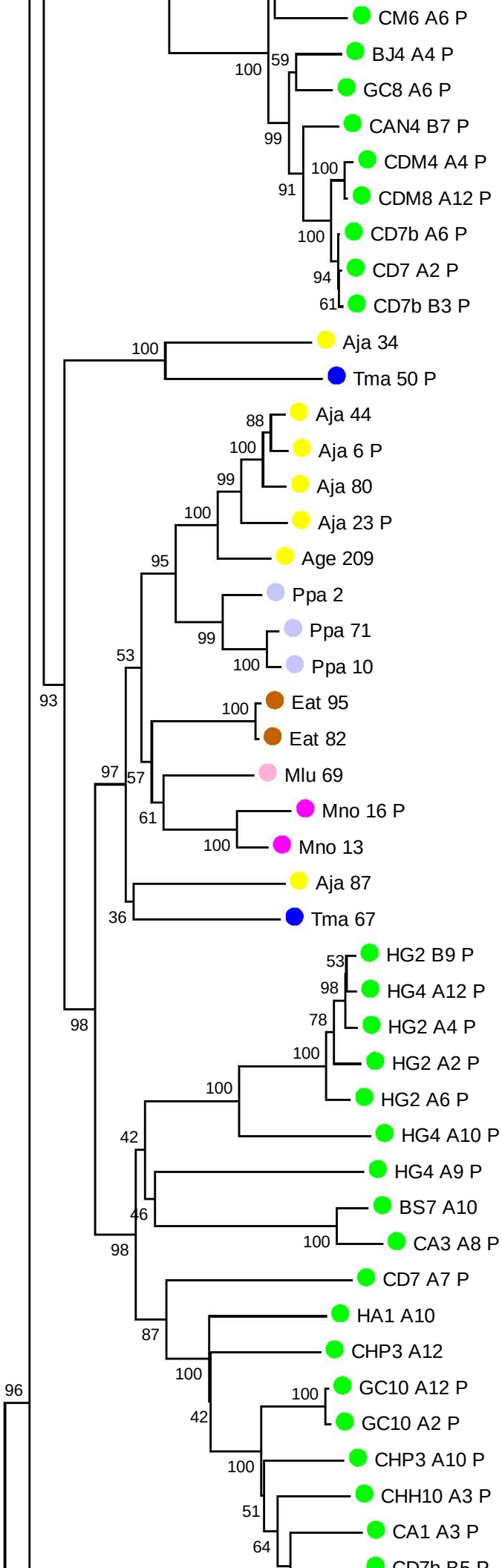


Figure 3 (cont.)

Back to overview

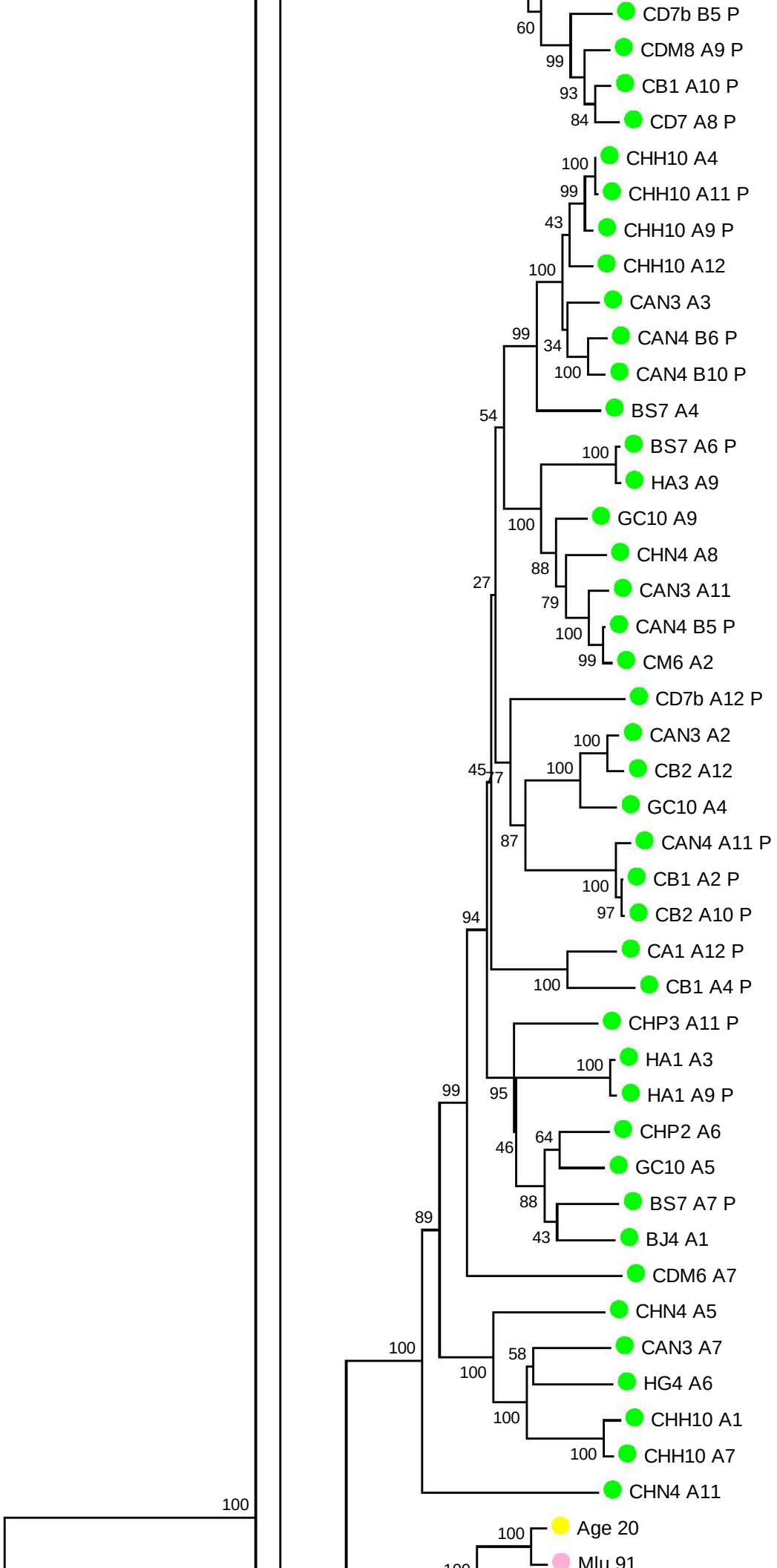


Figure 3 (cont.)

Back to overview

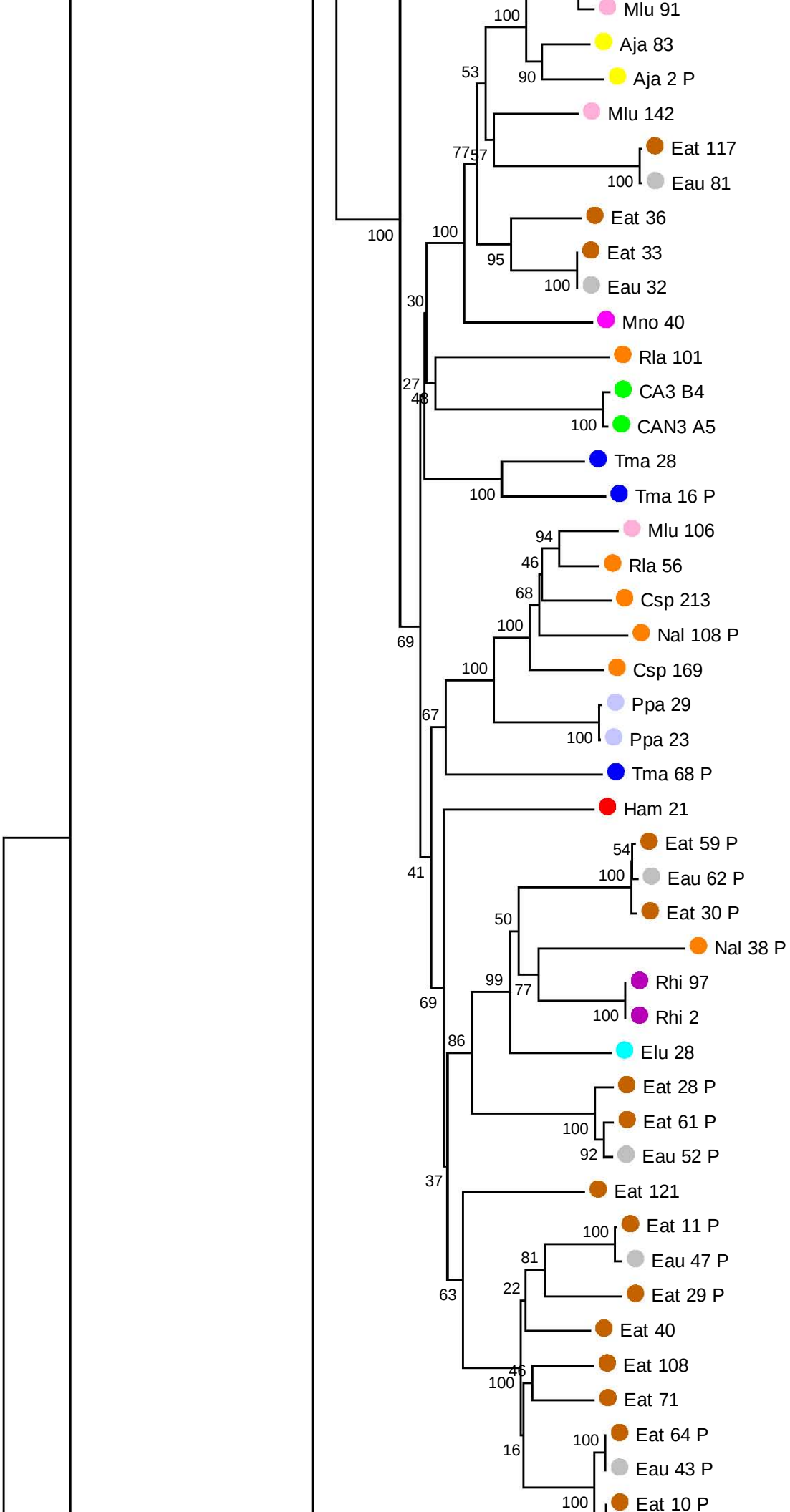


Figure 3 (cont.)

Back to overview

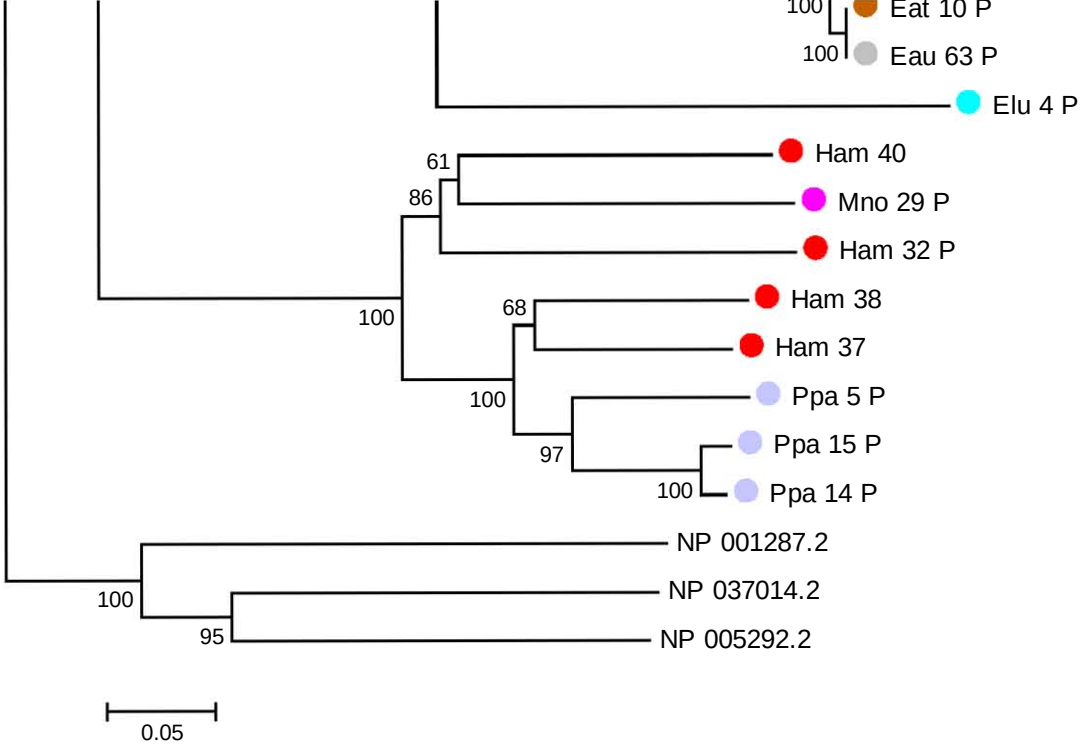


Figure 3 (cont.)

Back to overview